

**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**

Von der Fakultät für Ingenieurwissenschaften,  
Abteilung Informatik und Angewandte Kognitionswissenschaft  
der Universität Duisburg-Essen

zur Erlangung des akademischen Grades

Doktor der Philosophie (Dr. phil.)

genehmigte kumulative Dissertation

von

Jan Philipp Kluck  
aus  
Duisburg

1. Gutachterin: Prof. Dr. Nicole Krämer
2. Gutachter: Prof. Dr. Gerhard Vowe

Tag der mündlichen Prüfung: 23.09.2021



*This work is dedicated to my mother, Charlotte Kluck.  
Thank you for everything.*



## Acknowledgments

Writing this dissertation would not have been possible without the support of many wonderful people. Therefore, I would like to take this opportunity to thank all the people who have accompanied me on this challenging path. First of all, however, I would like to acknowledge the Ministry of Culture and Science of the German State of North-Rhine-Westphalia for funding this dissertation project within the Digital Society research program.

My deepest thanks go to Prof. Dr. Nicole Krämer for encouraging me to write this dissertation. Thank you, Nicole, for all the constructive feedback, for sharing your expertise, for all the support, and for often believing in me more than I did myself. Without you, this thesis would not have been possible. I am also deeply grateful to Prof. Dr. Gerhard Vowe for all the valuable and inspiring discussions about online communication over the last years.

Further, I would like to express my sincere gratitude to my project partner Marike Bormann, who accompanied me from the beginning to the end of the dissertation. Thank you for the great collaboration and all the civil talks about incivility. I would also like to thank all the other members of the Digital Society research program. Thank you all for the many wonderful moments at summer schools, conferences, and workshops.

I also owe special thanks to my (virtual) roommate Lea Frentzel-Beyme, who has been incredibly supportive over the past few weeks. Likewise, I would like to thank the whole SP team for the many inspiring conversations, the fun we had together, and the emotional support I received, which has contributed a lot to the success of this work. I will never forget this exciting time, and I am happy that I could spend it with you.

I am incredibly thankful for my family and friends, who always cheered me up and gave me the necessary breaks when I needed them. Thank you for always patiently participating in the studies and helping with the recruitment. Extraordinary thanks go to Malte, Mareike, and Peter Kluck for their tremendous support and encouragement. I am very glad that you have always been by my side. I would also like to thank Ruth and Heinz Diederichs for their continued interest in my work and all the support despite challenging times. Also, I want to thank Cosmo, who probably kept me company the most while writing this thesis.

Above all, from the bottom of my heart, I want to thank Elisabeth Wenzgol. Thank you, Ella, for your patience, for taking care of me, for always encouraging me, and for believing in me. Without you, I would not have been able to submit this work. Thank you for carrying me across the finish line.



## Annotation of the papers contained in the cumulus

### Article I

Kluck, J. P., & Krämer, N. C. (2020). It's the aggression, stupid! An examination of commenters' motives for using incivility in online political discussions. In *Proceedings of the 11th International Conference on Social Media and Society*, 164–173. <https://doi.org/10.1145/3400806.3400826>

### Article II

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.

### Article III

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>

### Article IV

Kluck, J. P., & Krämer, N. C. (2020). *Appraising Uncivil Comments in Online Political Discussions: How do preceding incivility and the sender's stance affect the processing of an uncivil comment?* Manuscript submitted for publication (Revise and resubmit).



## Zusammenfassung

Das Internet hat unser Leben in vielerlei Hinsicht verändert. Auch der politische Diskurs ist in großen Teilen in den digitalen Raum gerückt. Obwohl das Internet viele neue Möglichkeiten für politische Partizipation bietet, hat der Online-Diskurs auch seine Schattenseiten. Vor allem inziviles Kommunikationsverhalten in Online-Diskussionen hat viel Aufmerksamkeit in der Öffentlichkeit und in der Forschung erlangt. Die Folgen von Inzivilität sind allerdings schwer abzuschätzen, da sich inziviles Kommunikationsverhalten in sehr unterschiedlicher Form manifestieren kann und es sehr subjektiv ist, welche Diskussionsbeiträge als inzivil betrachtet werden. Deshalb ist es wichtig Wege zu finden, um Inzivilität auf einer perzeptuellen Ebene zu untersuchen. Letztendlich ist Inzivilität nicht nur das Merkmal einer Nachricht, sondern Menschen machen sich vermutlich auch ein Bild von der Person hinter der Nachricht. Ziele der vorliegenden Dissertation ist es daher einen Ansatz zu entwickeln, wie Inzivilität aus einer sozialen Perspektive beleuchtet werden kann. Nur so können wir verstehen, warum Inzivilität negative Folgen für die Dynamiken einer Diskussion und auch für die Rezipierenden selbst haben kann. Auch kann auf diese Weise exploriert werden, ob verschiedene Formen von Inzivilität unterschiedliche Auswirkungen auf die soziale Perzeption haben. Da Attributionstheorien erklären, wie Menschen das Verhalten anderer interpretieren, baut die vorliegende Dissertation insbesondere auf diesem Ansatz auf und integriert diesen in einen breiteren theoretischen Rahmen.

In Studie I wurde zunächst anhand qualitativer und quantitativer Daten untersucht, welche Motive Menschen haben, inzivile Nachrichten zu schreiben und wie diese sich zwischen verschiedenen Typen von Inzivilität unterscheiden. Die Ergebnisse zeigten, dass insbesondere zwei Arten von Diskussionsmotiven die Nutzung von Inzivilität begünstigen: aggressive Diskussionsmotive und das Motiv vorangegangene Normverletzungen zu konterkarieren. Allerdings führte nur das Vorhandensein von aggressiven Diskussionsmotiven zu häufigem inzivilen Kommunikationsverhalten.

Studie II hat anschließend auf der Grundlage von Annahmen über kooperative Kommunikation und Attributionsprozesse beleuchtet in welchem Maße verschiedene Typen von Inzivilität die Bewertung des kommunikativen Verhaltens beeinflussen. Dazu wurde für fünf verschiedene Typen von Inzivilität untersucht, wie sich die Attributionen der Rezipierenden über (1) Diskussionsabsichten, (2) tolerierbare politische Ansichten und (3) Diskussionsfähigkeiten des Absenders unterscheiden. Tatsächlich zeigte sich, dass hier signifikante Unterschiede gefunden werden können, die es erlauben Inzivilität auf einer perzeptuellen Ebene zu kategorisieren. Vor allem respektloses Diskussionsverhalten hat

negative Attributionen auf allen drei Ebenen hervorgerufen. Allerdings konnte nur im geringen Maße festgestellt werden, dass die verschiedenen Typen von Inzivilität das Diskussionsverhalten der Rezipierenden negativ beeinflussen.

Da insbesondere respektloses Diskussionsverhalten negative Attributionen hinsichtlich kooperativer Kommunikation ausgelöst hatte, wurde in Studie III und IV untersucht, inwieweit die Attribution von aggressiven Diskussionsmotiven eine Rolle für die Verarbeitung inziviler Kommentare spielt. Es zeigte sich, dass Menschen inzivilem Diskussionsverhalten in der Form von Respektlosigkeit ein hohes Maß an aggressiven Diskussionsmotiven zuschreiben. In Studie III zeigte sich zudem, dass dieses Muster in einer generellen negativen Bewertung des Senders resultierte. Dies führte wiederum dazu, dass Menschen weniger gewillt waren an der Diskussion teilzunehmen.

In Studie IV konnte zusätzlich gezeigt werden, dass die Zuschreibung von aggressiven Diskussionsmotiven die eigenen aggressiven Antwortintentionen der Rezipierenden erhöht hat. Nur wenn sie solche Intentionen hatten, war es wahrscheinlicher, dass sie selbst in einer inzivilen Art und Weise geantwortet haben. Ob sie solche Intentionen entwickelten, war auch von ihren positiven und negativen Emotionen abhängig.

Sowohl in Studie III als auch in Studie IV wurde zudem untersucht, wie die Verarbeitung von Inzivilität durch gegensätzliche Positionen beeinflusst wird. In beiden Studien zeigte sich, dass Menschen bei Person mit einer entgegengesetzten Meinung negativere Attributionen vornahmen als bei gleichgesinnten Personen. Allerdings konnte in Studie III festgestellt werden, dass dieser Effekt unabhängig vom Vorhandensein von Inzivilität auftrat.

Durch die Ergebnisse der vier Studien konnte ein Modell entwickelt werden, das beschreibt, wie Menschen Inzivilität von einer sozialen Perspektive verarbeiten und wie dies ihr eigenes Diskussionsverhalten beeinflussen kann. Die Annahme, dass Inzivilität auch auf einer sozialen Ebene verarbeitet wird, konnte so gestützt werden. Auch zeigte sich, dass insbesondere Attributionstheorien ein vielversprechender Ansatz sind, sich dieser Perspektive weiter zu nähern.

### **Abstract**

The Internet has changed our lives in many ways. Even the political discourse has shifted to a large extent into the digital space. Although the Internet offers many new opportunities for political participation, online discourse also has its dark sides. In particular, uncivil communication behavior in online discussions has gained notable public and scholarly attention. However, the consequences of incivility are challenging to assess, as it can take many forms, and it is highly subjective which discussion posts are considered uncivil. Therefore, it is essential to find ways to study incivility on a perceptual level. After all, incivility is not only a comment feature, but people are also likely to form an image of the person behind the message. Thus, this dissertation aims to develop an approach to examine incivility from the perspective of social perception. Only in this way can we eventually understand why incivility can have negative consequences for the dynamics of a discussion and the recipients themselves. Also, this approach allows exploring whether different types of incivility have distinct effects on social perception. Since attribution theories explain how people interpret the behavior of others, this dissertation builds particularly on this approach and integrates it into a broader theoretical framework.

First, Study I used qualitative and quantitative data to examine people's motives for writing uncivil messages and how these motives differ between distinct types of incivility. The results showed that two kinds of discussion motives facilitated the use of incivility in particular: aggressive discussion motives and the motive to counter previous norm violations. However, only the presence of aggressive discussion motives has led to frequent uncivil communication behavior.

Subsequently, Study II used assumptions about cooperative communication and attribution processes to explore the extent to which different types of incivility influence the evaluation of communicative behavior. For this purpose, we examined how recipients' attributions differ in (1) discussion intentions, (2) tolerable political views, and (3) discussion skills of the sender for five different types of incivility. Indeed, it turned out that significant differences can be found, which allow categorizing incivility on a perceptual level. Especially disrespectful discussion behavior elicited negative attributions on all three levels. However, the different types of incivility only slightly influenced the discussion behavior of the recipients.

Because particularly disrespectful discussion behavior had elicited negative attributions regarding cooperative communication, Studies III and IV examined the extent to

which the attribution of aggressive discussion motives plays a role in the processing of uncivil discussion comments. It was found that people attribute comparatively high levels of aggressive discussion motives to uncivil discussion behavior in the form of disrespect. Study III also showed that this pattern resulted in a generally negative evaluation of the sender. This, in turn, resulted in people being less willing to participate in the discussion.

Study IV additionally showed that the attribution of aggressive discussion motives increased the recipients' aggressive response intentions. Only if they had such intentions, they were more likely to respond in an uncivil manner themselves. Whether they developed such intentions was also dependent on their positive and negative emotions.

Both Study III and Study IV also examined how the processing of incivility was affected by opposing positions. In both studies, it was found that people made more negative attributions to a person with an opposing opinion than to like-minded people. However, Study III found that this effect occurred regardless of the presence of incivility.

Through the results of the four studies, a model has been developed that describes how people process incivility from a social perspective and how this may influence their discussion behavior. The assumption that incivility is also processed on a social level could thus be supported. It was also demonstrated that particularly attribution theories are a promising approach to approach this perspective further.

## Table of contents

<b>I Introduction.....</b>	<b>1</b>
<b>II Challenges for incivility research .....</b>	<b>3</b>
1 Incivility, a threat to (online) democracy? .....	3
2 The different facets of incivility.....	6
2.1 Definition and operationalization .....	6
2.2 Perception and measurement.....	9
3 New perspectives on (online) incivility – incivility as a social act .....	11
<b>III Empirical approach.....</b>	<b>12</b>
4 Incivility in online political discussions: A lack of cooperative communication, aggression, or both? .....	12
5 What drives incivility? .....	15
5.1 Motivation for uncivil discussion behavior .....	17
5.2 Person factors.....	18
5.3 First study .....	19
5.3.1 <i>Article 1: It's the aggression, stupid! An examination of commenters' motives for using incivility in online political discussions. (Kluck &amp; Krämer, 2020)</i> .....	20
5.3.2 <i>Conclusions from Study I and subsequent steps</i> .....	23
6 A lack of perceived cooperation? Appraising incivility in a social environment .....	25
6.1 Making sense out of others' (uncivil) behavior – the role of attribution theories.....	26
6.2 Second study .....	28
6.2.1 <i>Article 2: Unwilling, intolerable views, or incapable? – How different types of incivility can be distinguished based on attributions about cooperative communication (Kluck, Bormann, Rieß, &amp; Krämer, 2021)</i> .....	29
6.2.2 <i>Conclusions from Study II and next steps</i> .....	32
7 The role of attributed aggressive intent: Incivility within the general aggression model .....	34
7.1 The role of others' stance.....	37
7.2 The role of discussion norms.....	39
7.3 Third and fourth study.....	40
7.3.1 <i>Article 3: "What an idiot!" – How the appraisal of the writer of an uncivil comment impacts discussion behavior (Kluck &amp; Krämer, 2021)</i> .....	41
7.3.2 <i>Conclusions from Study III and purposes of Study IV</i> .....	43
7.3.3 <i>Article 4: Appraising Uncivil Comments in Online Political Discussions: How do preceding incivility and sender's stance affect the processing of an uncivil comment? (Kluck &amp; Krämer, 2020)</i> .....	45
7.3.4 <i>Conclusions from Study IV</i> .....	48
<b>IV Discussion .....</b>	<b>50</b>
8 General conclusions from the empirical studies .....	50

8.1 Incivility as a lack of cooperative communication.....	50
8.2 Incivility as aggressive communication .....	52
8.3 Combined perspectives .....	54
8.3.1 <i>The role of personality</i> .....	56
8.3.2 <i>The role of comment's stance</i> .....	57
9 Theoretical implications .....	57
9.1 Modeling the social processing of incivility.....	58
9.2 New opportunities to categorize incivility .....	61
9.3 The utility of attribution theories.....	62
10 Practical implications.....	63
10.1 Measuring incivility.....	63
10.2 Interventions .....	64
11 Limitations and future studies.....	65
12 Conclusion .....	66
<b>V References .....</b>	<b>68</b>
<b>VI Research articles .....</b>	<b>83</b>

**List of figures**

*Figure 1.* Schematic overview of the findings revealed by Study I. The dots indicate that the examined types of incivility are not exhaustive. .... 24

*Figure 2.* Visual categorization of the different types of incivility based on recipients' attributions. The illustration is adopted from the submitted manuscript..... 31

*Figure 3.* Visualization of the findings of Study II. Misleading information is crossed out, as it was not found that it influences attributions about cooperativeness. Moreover, it was not found that the usage of specific types of incivility is affected by the attribution process. However, recipients of topic deviation were slightly more likely to engage in further deviating behavior. In that cases, the norm violation itself probably has a direct contagious effect. The dots indicate that the examined types of incivility and findings are not exhaustive. .... 33

*Figure 4.* Illustration of the findings from Study III and Study IV. Preceding incivility is crossed out as no effects were found for this factor. Attributions influence whether people respond at all. If people respond, aggressive answer intentions determine whether people respond in an uncivil way. The arrow with the dotted line indicates the hypothetical reciprocal effect between attributions and emotions as well as cognitions..... 49

*Figure 5.* Model describing the social processing of incivility. Dotted lines indicate theoretical assumptions that have not been empirically investigated. The arrow pointing from incivility back to incivility indicates that certain types can facilitate further uncivil communication by themselves (e.g., deviating behavior). .... 60



## I Introduction

Conversations between citizens about political issues have been considered the “soul of democracy” (Kim et al., 1999, p. 362). As the Internet provides relatively new opportunities to participate in such conversations by globally connecting individuals, it has a certain potential to promote a democratic discourse. Primarily, social media platforms provide various spaces in which individuals can discuss politics and societally relevant issues. On social network sites like *Facebook* and *Twitter* and video-sharing platforms like *YouTube*, citizens can easily and informally discuss political issues (Boulianne, 2015; Halpern & Gibbs, 2013; Valenzuela et al., 2018; Valera-Ordaz, 2019). Likewise, news outlets provide comment spaces on their websites and become an integral part of social media platforms. That allows a fast distribution of news and the facilitation of related discussions (Kümpel et al., 2015; Ziegele et al., 2014). People can participate in conversations at relatively low costs on social media platforms by writing short comments, liking content, or sharing content. Moreover, people do not have to seek out (political) discussions proactively but are often incidentally exposed to discussions that emerged in their social media feeds and timelines. These dynamics have indeed been found to facilitate recipients’ online participation (Valeriani & Vaccari, 2016).

With rapidly changing opportunities to participate in public debates, concerns have grown that the share of uncivil communication has noticeably increased (Coe et al., 2014; Sobieraj & Berry, 2011). Such dynamics are usually considered a challenge for online political communication since incivility can have a range of adverse effects. For instance, it was found that discussion incivility increases negative emotions (e.g., Chen & Lu, 2017; Gervais, 2015, 2017) or can decrease recipients' open-mindedness towards other opinions (Hwang et al., 2018). However, it is difficult to assess the actual impact of incivility on political discourse, as there is no unique definition of this construct. Some scholars conceptualized incivility as disrespectful discussion behavior (e.g., Coe et al., 2014), whereas other scholars argue that incivility is a violation of democratic norms (e.g., Papacharissi, 2004). By now, incivility is primarily seen as a multidimensional concept that can take different forms (e.g., Chen, 2017; Stryker et al., 2016). Yet, little effort has been made to derive a common theoretical fundament that encompasses all types of incivility. One of the biggest challenges for research in this field is that incivility is a subjective concept because it highly depends on recipients' interpretation of the message, whether it is perceived as uncivil or not (Chen et al., 2019; Herbst, 2010). Therefore, incivility cannot be defined simply by the occurrence of specific message characteristics.

To close this gap, Bormann and colleagues (2021) recently introduced a novel approach to incivility. By drawing on distinct theoretical frameworks addressing cooperation, the authors conceptualize incivility as violations of different cooperative communication norms that are disapproved of by recipients. In other words, they regard online communication to be uncivil if it disappoints recipients' expectations for cooperative communication. In this way, Bormann and colleagues (2021) strengthen a common theoretical ground that allows categorizing different types of incivility based on the nature of the norm violation. Within this approach, a crucial aspect is highlighted: Incivility is not a mere message characteristic.

For successful cooperative actions, people need to use their unique expertise to read others' minds (Tomasello et al., 2005). In other words, they must infer what the goals of others are and what capabilities they have to achieve those goals. From this perspective, incivility is more than the usage of certain words. As online discussions are characterized by social exchange, incivility should be considered a social act rather than a message feature. So far, however, comparatively little empirical evidence exists for this social perspective. Therefore, the current doctoral thesis aimed to strengthen this approach by providing insight into how people socially process incivility in the online political discourse. At its very heart, *attribution theory* has been assumed to provide a vital framework to understand how people process incivility from social perception. Although there is no unique attribution theory (see Malle & Korman, 2013 for an overview of the different meanings), all approaches in this field have in common that they explain how people interpret the behavior of others. For behavior that is performed intentionally, the emphasis is on the attribution of behavioral motives within attribution theories (Jones & Davis, 1965; Malle, 1999; Malle & Korman, 2013; Reeder, 2009; Reeder & Brewer, 1979; Weiner, 2000). Therefore, special attention was paid to this aspect.

Consequently, the primary purpose of the dissertation was to examine which motives people have when behaving in an uncivil manner and which attributions recipients make about the senders of incivility. Based on the novel approach of Bormann and colleagues (2021), it was of particular interest whether (perceived) aggressive motives or a (perceived) lack of cooperation cause the undesirable consequences of incivility. By combining the attribution theory approach with assumptions about cooperative communication (Allwood et al., 2000; Bormann et al., 2021; Grice, 1975) and the evolvement of human aggression (Anderson & Bushman, 2002), an empirically based model was successively developed to approach incivility from the perspective of social perception.

## II Challenges for incivility research

### 1 Incivility, a threat to (online) democracy?

There is an ongoing debate about the actual contribution of online discussions to democracy (Coleman & Moss, 2012; Graham, 2015; Schäfer, 2011). Whereas some scholars focus on the benefits of online discussions for the public political exchange (e.g., Benkler, 2006; Brundidge, 2010), others have expressed concerns about the gain of such discussions for democratization (e.g., Dahlberg, 2001; Papacharissi, 2002). Although Papacharissi (2002) concedes that the Internet creates new spaces for political discussion and interpersonal contact, she emphasizes that the Internet will not automatically lead us to a democratic utopia by its technical features. She argues that it is rather crucial in which ways people use the Internet and which dynamics evolve in Internet-based public spaces.

Indeed, scientific research identified dynamics in online discussions that can be considered destructive to the democratic process. For instance, it was reasoned that social media platforms foster political polarization due to the way of disseminating political content (Settle, 2018). It was also found that anonymity in the online environment facilitates offensive comment behavior (Santana, 2014). However, a prevailing concern seems to be that uncivil communication behavior is increasing more and more (Coe et al., 2014; Sobieraj & Berry, 2011). Therefore, a discussion has arisen about the extent to which the Internet's potential to promote democratic processes is undermined by a decline of civility in online communication (e.g., Chen, 2017; Garton Ash, 2016; Halpern & Gibbs, 2013; Herbst, 2010; Papacharissi, 2004; Quandt, 2018; Rowe, 2015). Chen (2017) argues that incivility in online discussions could hinder *deliberative moments* – moments that “offer some support for the ideals of free debate, inclusiveness of viewpoints, and discussions across difference” (p. 176). On the other hand, scholars emphasize the merit of heated discussions for democracy and argue that a robust democracy could endure certain levels of uncivil communication (Garton Ash, 2016; Papacharissi, 2004).

Research attests that incivility is a common feature of (political) online discussions (Coe et al., 2014; Su et al., 2018). Likewise, for the US, recent survey data from the Pew Research Center indicate that adult Internet users increasingly experience uncivil behavior online – especially on social media. Whereas in 2014, 23% of adult Internet users personally experienced offensive name-calling, in 2020, 31% of users experienced such behavior. In the same period, users' experiences of purposeful embarrassment increased from 19% to 26%, and even incidents of physical threats have risen from 7% to 14%. Overall, in 2020, 41% of surveyed individuals reported having experienced some online harassment (Pew Research

Center; Vogels, 2021). Based on these dynamics, the German Government even enacted a law – the Network Enforcement Act (Netzwerkdurchsetzungsgesetz) – that obliges social media providers to delete uncivil comments as soon as someone reports content that violates the law (Federal Ministry of Justice and Consumer Protection, 2017).

Likewise, incivility in online discussions gained considerable scientific attention, and concerns have been expressed that incivility is indeed a serious threat to democratic discourse (Anderson et al., 2018; Ceron & Memoli, 2016; Han et al., 2018). So far, research has identified various consequences caused by incivility that may undermine democratic values. For instance, it was found that exposure to incivility fosters attitude polarization towards the discussion topic (Anderson et al., 2014) and increases perceptions of a polarized public (Hwang et al., 2014). Similarly, Goyanes and colleagues (2021) found that uncivil discussion comments can facilitate disengagement from others (i.e., unfriending) on social media, which can foster the avoidance of contentious political content. Consequently, this outcome might also boost polarization dynamics. Although such dynamics can also encourage citizens to participate politically (e.g., Borah, 2014), concerns are that polarization leads to a more intolerant and divided society (Stroud, 2010). It was also demonstrated that uncivil comments related to online news increase perceptions of a biased coverage (Anderson et al., 2018). As it was found that people holding such biased perceptions are more likely to speak out and to seek out differences in political conversations (Barnidge & Rojas, 2014), incivility has seemingly the potential to decrease diversity in online discussions. However, the probably most detrimental consequence of incivility in online discussions for the (online) democracy is that it can directly discourage citizens from participating in discussions. Han and Brazeal (2015) found that exposure to an uncivil discussion decreases individuals' willingness to participate in future online discussions. This seems especially true for individuals who try to avoid conflicts (Sydnor, 2019). Moreover, Muddiman and colleagues (2020) demonstrated that people are less likely to engage with online news that tends to be uncivil. In turn, it was found that individuals who frequently comment in discussions are less likely to do this in a civil and informational manner (Blom et al., 2014). From this perspective, incivility hinders democratic discourse by diminishing diversity of opinion and encouraging only certain people to discuss.

At this point, it is important to note that it was also found that incivility can have positive consequences on the democratic process. For instance, it was demonstrated that incivility can foster political engagement. Borah (2014) showed that incivility in the blogosphere increases the willingness to join an online petition. Likewise, Chen and Lu

(2017) found that people exposed to uncivil comments posted on a news story about abortion were more likely to engage in political activities (online and offline) regarding this issue. Additionally, it was found that incivility can encourage individuals to speak out when having a minority opinion through the way of emotions (Masullo et al., 2021). Although these effects of incivility appear favorable at first glance, the actual consequences of these effects have not yet been sufficiently researched. This said, when responding to incivility, people tend to react in an uncivil way as well (Chen & Lu, 2017; Gervais, 2015; Kwon & Gruzd, 2017; Ziegele et al., 2018). Thus, speaking out in an uncivil discussion might end up being rather destructive because aggression in comment sections generates further aggression (Gonçalves, 2018), and incivility is likely to encourage individuals to participate politically due to defensive motivations (Borah, 2014; Chen & Lu, 2017; Hwang et al., 2008). From this perspective, incivility will probably prevent deliberative moments, as Chen (2017) described, even when it encourages users to speak out. Although political engagement is an important cornerstone of our democracy, it is less valuable for democracy when it is accompanied by close-mindedness and polarized attitudes.

Taken together, there is evidence that incivility in online political discussions can have certain adverse effects – at least when these discussions take place in public. Nevertheless, scholarly voices against seeing incivility as inherently bad are getting louder (Chen et al., 2019). Already Papacharissi (2004) reasoned that even rude and heated discussions could benefit democracy as long as they promote democratic exchange. However, the adverse effects described above show that incivility can have undesirable consequences for the public online political discourse.

One major challenge for research on incivility is that it is often very subjective and context-dependent, whether a comment is uncivil or not. For that reason, (in)civility is considered “notoriously” difficult to conceptualize (Boyd, 2004; Coe et al., 2014; Muddiman et al., 2020). This had led to distinct approaches to defining and operationalizing incivility. Therefore, it is difficult to determine how incivility actually has adverse consequences and when it even might foster online political discussions. Likewise, research often neglected to differentiate between distinct types of incivility when examining its consequences. Usually, the effects of incivility were investigated by exposing participants to various uncivil comments, mixing different incivility types (e.g., Masullo et al., 2021; Hutchens et al., 2019). The following section discusses how incivility was defined, operationalized, and measured so far.

## 2 The different facets of incivility

In the broadest sense, incivility can be considered as norm violating communication (Jamieson & Hardy, 2012; Muddiman, 2017; Oz et al., 2017; Papacharissi, 2004; Rossini, 2019; Sydnor, 2018). However, which communicative behavior is compliant with norms and which behavior is norm violating is often context-dependent and can change over time (Chen et al., 2019; Herbst, 2010; Rossini, 2019). Therefore, the perception of civility and incivility is “very much in the eye of the beholder” (Herbst, 2010, p. 3). The target of incivility can be very different, too. For instance, incivility can be aimed at recipients themselves or against other discussants (Chen & Ng, 2016, 2017). Incivility can also be aimed either against people present in the discussion or against people who are not present (e.g., authors of a news article or politicians; Rowe, 2015; Su et al., 2018). Furthermore, it can be aimed at a whole (often marginalized) social group (Chen, Fadnis, et al., 2020; Papacharissi, 2004) or can be against institutions like news media, companies, or governments (Papacharissi, 2004; Rost et al., 2016).

Consequently, incivility in online political discussions can take many forms. However, although online incivility is far from being a new phenomenon, as mentioned above, there is an ongoing debate about its definition (Chen et al., 2019). In the next section, it will be discussed which definitions and operationalization exist so far and where are still shortcomings in the literature.

### 2.1 Definition and operationalization

Often, incivility is defined and operationalized as disrespectful and offensive comments towards other discussants (Chen & Lu, 2017; Coe et al., 2014; Gervais, 2014; Sobieraj & Berry, 2011; Sydnor, 2019). Coe and colleagues (2014) defined incivility as “features of discussion that convey an unnecessarily disrespectful tone toward the discussion forum, its participants, or its topics” (p. 660). Consequently, they operationalized incivility as comments that include name-calling, aspersion, lying accusations, vulgarity, and pejorative speech. In the past, these manifestations of incivility have often been conceptualized as uninhibited online communication and have been labeled as *flaming* (Lea et al., 1992). Other scholars argue that incivility is more than an impolite discussion tone. Papacharissi (2004) argues that a more significant distinction needs to be made between impoliteness and incivility. Whereas impoliteness does not necessarily decrease the democratic merit of a discussion, she argues that only behavior with a lasting negative impact on democracy can be considered uncivil. Therefore, she defined incivility as “disrespect for the collective traditions of democracy” that can be operationalized “as the set of behaviors that threaten democracy,

deny people their personal freedoms, and stereotype social groups” (p. 267). Indeed, people seem to differentiate between these forms of incivility. Muddiman (2017) introduced the concepts of personal-level incivility and public-level incivility. This differentiation comes close to the distinction made by Papacharissi (2004). Whereas personal-level incivility refers to violations of politeness norms, public-level incivility refers to violations of political and deliberative norms and, therefore, corresponds to Papacharissi's (2004) idea of true incivility. In her study, it has to be noted that the operationalization of public-level incivility was much broader than the one of Papacharissi (2004). For instance, in Muddiman's (2017) study, deceptive behavior (e.g., spreading rumors or accusations of lying) was also assigned to public-level incivility. In a more recent approach, Rossini (2022) distinguishes between incivility and intolerance. In accordance with previous suggestions (Coe et al., 2014), she conceptualizes incivility in terms of a disrespectful and impolite discussion tone that, however, does not necessarily lack justification. In contrast, she argues that political intolerance is inherently harmful to democracy as it threatens the rights of individuals or specific groups, signals moral disrespect, and fosters violence.

Therefore, the most important distinction seems to be the differentiation between the violation of interpersonal respect and politeness norms as well as the violation of democratic and deliberative norms. Although these approaches are beneficial and vital, there is still a need to classify uncivil discussion behavior further as incivility can take forms that go beyond this two-dimensionality. In this vein, Chen (2017) stated that incivility could be conceptualized as a “continuum of non-normative speech that ranges from impoliteness to virulent hate speech” (p. 84).

Another aspect is that the danger posed by the violation of politeness norms is often deemphasized (Papacharissi, 2004; Rossini, 2022), even though previous research clearly shows that people perceive disrespectful communication to be highly uncivil (Kenski et al., 2020; Muddiman, 2017; Stryker et al., 2016). Following previous findings (Han & Brazeal, 2015), Rossini (2021) also found that incivility focusing on violations of respect and politeness norms is less likely to get replies in online discussions. Thus, the threat to democracy by this type of incivility must not be underestimated as it can prevent an open political discourse by discouraging individuals from participating.

Chen et al. (2019) noted that it might be helpful to focus on the actual outcomes of incivility rather than focus on the usage of certain words or phrases. Indeed, in this manner, new ways could be found to categorize distinct types of incivility and examine which types of incivility should be encountered and which types are less harmful to online discussions.

However, so far, little effort has been taken to categorize incivility based on its consequences. A major problem is that the distinct definitions and conceptualizations of incivility are seldomly theory-driven but usually depend on the conceptual differences in its manifestation.

Bormann and colleagues (2021) address this shortcoming by proposing a theory-based typology of incivility. The typology categorizes different types of uncivil communication based on their common ground: The authors argue that incivility in political discussions has its origins in the violation of individuals' expectations of others to adhere to norms of cooperative communication. It is emphasized that cooperative communication does not exclude dissent between different parties but can even foster the discourse by allowing cultivated conflicts (see also Garton Ash, 2016, p. 148 ff.). In other words, to have merit for the democratic discourse, political discussions do not have to be consensus-oriented, and people do not always have to agree with each other. A heated debate is not necessarily uncooperative in this vein – even if it is not overly polite. Therefore, this perspective meets Papacharissi's (2004) concerns that heated disagreement must not be equated with incivility. However, on that note, a certain degree of cooperative communication is needed so that deliberative moments can occur, as proposed by Chen (2017). Consequently, Bormann and colleagues (2021) conclude that incivility can be defined as communicative acts in public (online and offline) political discussions that violate certain norms of cooperative communication and is severely disapproved by the debate's participants. Then, different types of incivility can be conceptualized as violations against distinct norms of cooperative communication that are expected by discussants in political (online) discussions. Drawing on this assumption, the authors derived five generic communication norms, which, if violated, can have a decisive negative impact on cooperative communication.

Based on Grice's (1975) cooperation principle, Bormann and colleagues (2021) deduced norms they assume individuals expect others to adhere to when communicating. First, the authors propose an *information norm* and a *modality norm*: The information norm refers to Grice's (1975) communication maxims of *quantity* (e.g., “Make your contribution as informative as is required”) and *quality* (e.g., “Do not say what you believe to be false”) whereas the modality norm refers to the maxim of *manner* (e.g., “Avoid ambiguity”) (p. 45 – 46). Thus, from this perspective, behavior like spreading disinformation or using misleading exaggerations can be considered violations of the information norm. In contrast, although more seldomly addressed in the literature of incivility, ambiguous communication behavior like being ironic or sarcastic can be considered as violations of the modality norm. Moreover, Bormann and colleagues (2021) argue that cooperative communication can only succeed

when it is connected and consistent. Therefore, they propose a *process norm* that postulates that ignoring others' discussion contributions or deviating from the discussion topic can be regarded as significant norm violations of cooperative communication. Although not explicitly stated by Bormann and colleagues (2021), this norm also corresponds to Grice's (1975) *relation maxim* (i.e., "Be relevant"). However, Grice's (1975) leaves largely open what that norm includes exactly.

Nonetheless, ignoring others' discussion contributions or deviating from the discussion topic can be regarded as significant violations of the process norm. Corresponding to more common types of incivility (e.g., Coe et al., 2014), Bormann and colleagues (2021) also propose a *relation norm* that states that cooperative communication should be respectful. Hence, being offensive and disrespectful to other discussants is a violation of this norm. Finally, the authors propose a *context norm* stating that communication should consider the communicative context to enable cooperation. Regarding (online) political discussions, the authors argue that public communication should meet the principles of liberal democracies. Thus, incivility in the form of democratic threatening behavior as operationalized by Papacharissi (2004) can be allocated to violations against the context norm.

Although Bormann and colleagues (2021) take a strong perspective from communication science and sociology, this framework is also helpful to approach incivility from a psychological perspective. First, the core of the approach roots in evolutionary psychology (Tomasello, 2008, 2019; Tomasello et al., 2005), and reasoning from social psychology (e.g., Cialdini et al., 1991; Kerr & Kaufman-Gilliland, 1994) is broadly integrated. Second, whereas incivility is often treated as a feature of a message (e.g., Coe et al., 2014; Muddiman, 2017; Papacharissi, 2002; Rossini, 2021), this framework puts the focus on what incivility actually is – a social act. In this way, Bormann and colleagues (2021) acknowledge the subjective nature of incivility by conceptualizing it as a perceptual construct. Yet, empirical validation of the framework is missing. So far, individuals' assessment of incivility itself has received relatively little attention. As diverse as the definitions and conceptualizations are, as diverse are the approaches to measure perceived incivility.

## 2.2 Perception and measurement

As incivility can be considered a perceptual construct (Bormann et al., 2021), it is essential to explain what differences exist in the individual perception for understanding which effects different incivility types have on its recipients. However, this aspect has been largely neglected in the literature. Perceptions of incivility primarily served as a manipulation check (e.g., Anderson et al., 2018; Hutchens et al., 2019; Hwang et al., 2014) or to pre-test

stimulus material (Chen & Lu, 2017; Chen & Ng, 2016). Moreover, only a few studies examined differences in the perception of uncivil comments between distinct types of incivility. For instance, Stryker et al. (2016) investigated lay people's consensus about the perception of incivility among different communicative acts. The authors found that there is high agreement about the perceptions of incivility and, based on individuals' assessments, they deductively identified three distinct dimensions of incivility: *utterance incivility* (e.g., name-calling, vulgarity), *discursive incivility* (e.g., interruption, refusal to listen), and *deception* (e.g., misleading, no evidence). Muddiman (2017) examined whether her proposed distinction between personal-level incivility and public-level incivility can be validated on a perceptual level. She found that people differentiate between the two concepts and that – against Papacharissi's (2004) reasoning – people consider personal-level incivility as more uncivil than public-level incivility. Likewise, Kenski and colleagues (2020) examined whether people perceive distinct types of incivility as operationalized by Coe et al. (2014; i.e., name-calling, aspersion, lying accusations, vulgarity, pejorative of speech) differently. Results demonstrated that individuals perceived name-calling and vulgarity as more uncivil than the other types of incivility.

Thus, overall, research indicates that there is variance in the perception of incivility based on its manifestation. However, although these studies are critical first steps to highlighting the perceptual character of incivility, it remains inconclusive how the differences in perception affect individuals. First, the mere perception of whether a communicative behavior is uncivil or not says little about the actual impact of the behavior. For instance, people can perceive something as uncivil while at the same time considering incivility as justified. Second, the reliability of the employed measures of perceived incivility is limited. In the studies of Stryker et al. (2016) and Muddiman (2017), perceptions of incivility were measured on a single item scale, asking participants how uncivil particular behavior or comments are. Kenski and colleagues (2020) measured perceived incivility with five items corresponding to the definition of incivility from Coe et al. (2014): *civil/uncivil*, *polite/rude*, *necessary/unnecessary*, and *respectful/disrespectful*. Although this approach is more comprehensive than a single-item measure, it focuses on a specific conceptualization of incivility, lacking a theoretical foundation. Thus, since incivility is a very complex phenomenon and in none of the studies individuals were exposed to a definition or conceptualization of incivility, it remains inconclusive what actually has been measured. Likewise, these studies hardly investigated what consequences follow from the perception of

incivility. Therefore, new ways must be found to understand how people appraise incivility and how this influences them on a behavioral level.

### **3 New perspectives on (online) incivility – incivility as a social act**

In sum, it can be noted that research on incivility so far lacks a theoretical foundation in different ways. The distinction between different types has been relatively unsystematic and lacks comprehensive theoretical derivation with a common ground. Therefore, the approach introduced by Bormann and colleagues (2021) is to be applauded as it does not only help to differentiate distinct types of incivility based on a more robust theoretical fundament but also highlights the need for a deeper understanding of how people appraise incivility on a perceptual level. Nevertheless, here, the approach of Bormann and colleagues (2021) is also limited since it mainly focuses on the systematization of different types of incivility. Moreover, their proposed typology and definition of incivility still lack empirical validation. In particular, how individuals process distinct uncivil messages and which roles perceptions of cooperation and conflict play have not been investigated sufficiently. Only when we understand why people appraise certain messages to be uncivil, we can predict which reactions will follow and, therefore, assess incivility's consequences.

From this perspective, new opportunities to measure and operationalize incivility can be derived that consider the context and individuals' subjective interpretation. Especially, the role of social perception has been neglected in this context. Yet, the appraisals of uncivil commenters are probably a more reliable measure of uncivil behavior than the mere assessment of a message as it considers the context and the relationship between communicators. Although this point of view is not entirely new (see O'Sullivan & Flanagin, 2003), it gained little further attention within the literature on online incivility.

This perspective is critical in the online context: Due to the audiovisual anonymity in online discussions, people have to make inferences about other commenters primarily based on the content of their messages. The way in which an opinion is expressed has, therefore, particular importance for recipients' interpretation of the message in such settings. Also, the number of individuals reading in public online discussions is probably higher than in offline discussions, as people can read and participate in an online discussion with little cost. Based on these factors, uncivil messages might have a more substantial impact in online discussions than in offline discussions. Consequently, a research model will be derived that will be examined in the context of online discussions but is not necessarily limited to the online environment.

### III Empirical approach

#### 4 Incivility in online political discussions: A lack of cooperative communication, aggression, or both?

A fundamental framework on which Bormann and colleagues (2021) draw stems from evolutionary psychology. Tomasello's (2008, 2009, 2019) work highlights that cooperative communication is basically built on shared goals and jointly coordinated to reach the goal. Here, Tomasello and colleagues (2005) describe well what is needed for cooperative activities (the authors originally used the term collaboration) from a person-centered perspective: Joint intentions result from shared goals and coordinated action plans on how to achieve the respective purpose. Likewise, interaction partners need to give attention to each other to understand each others' intentions. Tomasello and colleagues (2005) argue that understanding each others' intentions and forming joint intentions is a unique ability of humans that distinguishes them from animals. Although this describes very basic human social behaviors, it can also be well applied to (online) communication. Only when people share a communication goal, understand each others' intentions, and basically coordinate the communication can the communication outcome be constructive. However, the process described by Tomasello and colleagues (2005) only describes the basic mechanisms of how cooperation (or collaboration) works. In this way, it only rudimentarily considers which factors can affect cooperation negatively or even lead to a failure of cooperation. Concerning communicational acts, Allwood and colleagues (2000) provide a more nuanced view on cooperative behavior. In accordance with the conceptualization of shared intentionality by Tomasello et al. (2005), Allwood and colleagues (2000) argue that cooperation requires at least that involved agents consider each other cognitively and that the agents have joint purposes.

However, the authors state that the involved agents must not automatically be aware of the latter requirement. This notion is an important supplement. Especially in communication, people are probably not always aware of the communication goal. Rather, they have implicit expectations of how communication should proceed to be successful (Grice, 1975). For online discussions, this means that people do not necessarily be aware of each others' purposes. Nevertheless, each agent must have the same purpose for cooperative communication on a basic level to develop. For instance, discussing a particular political topic would be the minimum of shared purpose in online political discussions. Thus, people neither must share the same opinion on the topic nor a prerequisite for cooperative communication that

individuals have the shared purpose to reach an agreement. Quite the contrary, Allwood and colleagues (2000) state that purposes can be of antagonistic nature (e.g., hurting each other). Even when this purpose can be induced by only one side, on a basic level, this purpose will probably be adopted by the other side as long as the agents consider each other cognitively. Therefore, on this basic level, even uncivil communication can be cooperative to some degree.

As the cited requirements largely correspond to the basic mechanisms of shared intentionality proposed by Tomasello and colleagues (2005), Allwood and colleagues (2000) add further prerequisites that are needed for what they name *ideal cooperation*. First, they state that ideal cooperation requires not only mutually cognitive consideration but also *ethical consideration*. Basically, this means that agents should act in an ethical manner when interacting with others. That includes not forcing others, not trying to prevent others from pursuing their own goals, or not trying to prevent others from acting rationally. Moreover, Allwood and colleagues (2000) name *trust* as the last precondition for ideal cooperation. In this context, trust does not mean holistic trust in each other, but that the other agent(s) is willing and able to fulfill the other requirements.

When others do or cannot comply with the respective obligations, social sanctions may be the consequence (Allwood et al., 2000). In other words, when individuals violate norms of cooperation, others are likely to punish them. By punishing norm violators, Tomasello (2019, Chapter 9) states that individuals aim to reinforce social norms and restore their justice perceptions. Here, Tomasello (2019) emphasizes that aggression is the usual way to encounter perpetrators (p. 263). Noteworthy, as outlined above, the resulting behavior is not necessarily uncooperative if the involved persons adopt the same purpose of the communication and cognitively consider each other. Nevertheless, it probably does not meet the requirements for ideal cooperation as individuals do not ethically consider each other (e.g., preventing each other from reaching their respective goals).

Moreover, it is important to note that people do not seem to punish others because of the mere norm violation. Instead, the perceived intention appears to be crucial for individuals' reactions. Based on experiments with toddlers, Kachel et al. (2018) found that an intentional norm violation in a collaborative game is likely to arouse emotional protest (i.e., punishment). In contrast, when norms of collaboration were violated due to ignorance, toddlers were more likely to be engaged in teaching the roles than to protest the behavior.

In conclusion, whether communicative cooperation arises or not depends on individuals' intentions and the perceived intentions of others. Therefore, people must have a certain level

of cognitive awareness of each other. However, whether the cooperation proceeds ideally depends on individuals' degree of ethical considerations and their trust in each other. Especially when people have antagonistic intentions, these requirements are likely to fail. Thus, aggressive intent can prevent ideal cooperative communication, although basic cooperation can still emerge.

For incivility in online political discussions, this means that cooperation does not necessarily fail at all, but ideal cooperation is likely to be hampered. However, little is known about people's intentions when behaving uncivilly and which intentions people "read" within the other side. As the appraisal of incivility is largely subjective (Chen et al., 2019; Herbst, 2010), it is not given that peoples' actual intentions when behaving in an uncivil manner are correctly assessed by recipients. Thus, misperceptions of real and inferred intentions might hamper further cooperative communication. Especially when distinguishing between different types of incivility (see Bormann et al., 2021), differences may arise within and between individuals' intentions and the intentions recipients infer from the behavior.

To understand how individuals assess others' behavioral intentions, attribution theories are helpful. These theories address how people explain others' behavior and which behavioral intentions they infer from the behavior (Heider, 1958; Jones & Davis, 1965; Kelley, 1967; Malle, 1999; Reeder & Brewer, 1979; Weiner, 2000). Through the attributions recipients of incivility make, it can be examined whether they lose trust in the sender's cooperativeness and to which extent they perceive aggressiveness. Therefore, the overarching research questions of the current thesis are:

- *To which degree do people have cooperative and aggressive intentions when commenting in an uncivil manner in online political discussions?*
- *To what degree do people attribute cooperative and aggressive intentions to others who comment in an uncivil manner in online political discussions, and how do these attributions influence the recipients' assessment of the commenter?*
- *How do the attributed intentions and the sender's assessment of an uncivil message influence recipients' discussion behavior?*
- *How do these patterns differ between different types of incivility?*

Answering these questions has several merits. First, knowing how people assess the sender of an uncivil message helps us to understand how people process uncivil messages. We can only determine when incivility has adverse effects on recipients and if it should be

counteracted when we have this understanding. Second, assessment of the sender of an uncivil message, particularly the senders' intentions, is probably a more reliable way to measure perceived incivility as it captures incivility beyond the mere evaluation of certain words and phrases. Third, when understanding why different types of incivility have which outcomes, in supplement to the approach of Bormann et al. (2021), different types of incivility can be categorized by the consequences they cause.

From a theoretical perspective, it stands to reason that incivility roots in both a lack of cooperative communication and aggression. Tomasello (2019, Chapter 9) argues that perceptions of uncooperative behavior can trigger feelings of aggression. Therefore, uncooperative communication behavior can lead to aggressive communication. In turn, aggressive communication can become the new basic cooperative communication if communicators consider this as the new joint purpose (Allwood et al., 2000). However, so far, the literature on incivility fundamentally lacks insights into these mechanisms. Based on surveys and experimental designs, the current dissertation aims to close this gap by investigating what people drive to communicate in an uncivil manner and how recipients process such behavior.

As a first step of the thesis, it will be explored which motives people have when behaving in an uncivil manner to gain knowledge about the driving factors behind incivility and how these factors differ between distinct types of incivility.

## **5 What drives incivility?**

Based on the adverse effects of incivility, scientific research examined which factors cause and facilitate uncivil commenting behavior in the online environment. As communication on the Internet, especially on social media, has specific characteristics that fundamentally differ from face-to-face communication, research initially focused on how certain features of online discussion spaces influence the emergence of incivility. For instance, it has been argued that social media promotes impulsivity (Ott, 2017) and that impulsive usage of social media facilitates incivility (Turel & Qahri-Saremi, 2018). In this vein, it was also reasoned that the usage of mobile devices reinforces impulsive commenting behavior on social media and, therefore, facilitates certain types of uncivil and impolite commenting behavior, which – at least for the latter – could indeed be demonstrated (Groshek & Cutino, 2016).

Moreover, the visual anonymity in the online environment was considered the main driver for antinormative and disinhibited behavior for a long time (e.g., Dahlberg, 2001; Lea & Spears, 1991; Papacharissi, 2002; Rowe, 2015; Santana, 2014). The assumptions have been

that individuals experience “a loss of identity leading to an antisocial state” and “a lack of rationality” due to the absence of social cues in computer-mediated communication (Lea & Spears, 1991, p. 285). Noteworthy, recent research found that the perceived anonymity in online discussions is related to perceptions of uncivil online commenting, whereas the actual anonymity was no significant predictor (Van Duyn & Muddiman, 2022).

Following the model of social identity and deindividuation effects (SIDE), anonymity in online communication can also increase feelings of a salient social identity and, in this way, can also foster group normative behavior (Postmes et al., 1998, 2001). Drawing on this approach, Rösner and Krämer (2016) found that people have been more likely to comment in an uncivil way when peers commented in an uncivil manner before. They also demonstrated that anonymity had no direct effect on the uncivil commenting behavior but that people were more likely to conform to an uncivil norm in anonymous discussion settings than in non-anonymous discussions. These findings indicate that others’ behavior plays a vital role in uncivil online behavior. Accordingly, Frischlich and colleagues (2021) found that the strongest predictors for uncivil online behavior roots in the experiences with others’ communication behavior. Foremost, people who personally experienced online victimization were more likely to behave in an uncivil manner. Similarly, witnessing uncivil communication behavior increased the likelihood of uncivil online participation by individuals themselves. It must be noted that in this study, uncivil participation was not only operationalized by producing uncivil content but also by sharing or liking such content. Nevertheless, Koban and colleagues (2018) also found that perceived provocations positively predicted uncivil communication behavior.

Whereas these findings support the idea of the “social side” of incivility, little is known about the goals people pursue with such behavior. However, as people are “goal-oriented active communicators seeking to gratify felt needs” (Alonzo & Aiken, 2004, p. 207), the motivation for uncivil online behavior can tell a lot about the causes of incivility. Moreover, although a sender’s personality has been identified as a predictor for uncivil online behavior, it was hardly examined how different personality traits foster uncivil commenting. Therefore, the first study of the cumulus aimed to explore the motives people have for behaving uncivilly and which role their personality plays in this context.

This step has three merits: First, peoples’ actual motives can be put in relation to recipients’ appraisal of these motives. Second, different types of incivility could be categorized by differences in senders’ motives. Third, by exploring commenting motives for

different types of incivility, we can better generate assumptions about distinct effects on recipients, respectively.

### 5.1 Motivation for uncivil discussion behavior

Humans inherently strive for control over the physical and social environment to achieve desired goals (Heckhausen, 2000; Heckhausen & Heckhausen, 2018, Chapter 1; Ryan & Deci, 2000). When it comes to motivated actions, individuals have to coordinate their perceptions, emotions, cognitions, and skills to achieve the anticipated outcomes in the environment (Heckhausen, 2000; Heckhausen & Heckhausen, 2018, Chapter 1). Although this very basic perspective is highly characterized by evolutionary psychology, the idea can be transferred to the online environment. When people perform certain behavior online to achieve a specific desired goal, they will cause changes in the respective environment that others can observe. However, which goals people want to accomplish when behaving uncivilly have been largely neglected in research for a long time (see also Santana, 2014).

When examining individuals' motivation for media usage, researchers often draw on the *uses and gratification theory*. Originally developed for "traditional" media like television and radio, the theory postulates that people deliberately choose certain media content to receive certain (expected) gratifications like tension release or knowledge gain (Katz et al., 1974; Palmgreen & Rayburn, 1982). However, this approach has also been transferred to the context of the Internet (Papacharissi & Rubin, 2000; Ruggiero, 2018; Sundar & Limperos, 2013). Primarily, research on social media usage broadly draws on the uses and gratifications theory to explain users' behavior. This research includes the examination of users' motives for general usage of social networking sites like Facebook (e.g., Cheung et al., 2011; Smock et al., 2011) but also specific behavior on social media like photo and video sharing (Malik et al., 2016), engagement with social media content (Dolan et al., 2016), social bonding (Phua et al., 2017), or interpersonal communication (Eginli & Tas, 2018). Until now, little research addressed which motives people have when engaging in online commenting spaces and, more specifically, which gratifications they sought when behaving uncivilly.

Springer and colleagues (2015) conducted one of the few studies that examined what motivates people to write and read user comments in the context of online news articles. They found that motives related to social interaction and the desire to discuss were the main driver for commenting behavior. This suggests that when people comment in online spaces, they have a desire that has to be gratified by influencing the social environment. Regarding incivility, Alonzo and Aiken (2004) examined which motives individuals have to behave uncivilly online by commenting in an offensive and disrespectful manner (in this study

labeled as flaming). The authors found that such behavior is driven by the desire to pass the time, escape, relax, or desire entertainment. In the context of YouTube, other scholars found that people use this type of incivility to express anger, react to offensive behavior, or express disagreement and critique (Moor et al., 2010).

Although these early insights into the motivation for uncivil online behavior are beneficial, these studies have some shortcomings. First, in the study of Alonzo and Aiken (2004), participants were requested to assess motives that originally stem from research on television usage. Likewise, in the study of Moor and colleagues (2010), participants were exposed to a limited amount of preselected statements. Therefore, certain motives may have remained undetected. The authors also focused only on one specific type of incivility. As incivility can take many forms, the motives for distinct types of uncivil behavior probably differ as well. For instance, it is unlikely that people have the same motives for insulting others as they have for stating negative stereotypes. Furthermore, no study has explored individuals' motives for acting uncivilly in the context of online political discussions.

Hence, more research is needed on the motivation for uncivil online communication, especially in online political discussion. In this way, we can not only better understand individuals' intentions, but we can also identify contextual factors that promote uncivil discussion behavior. However, behavioral motives do not emerge independently but are closely related to an individual's person characteristics. Therefore, in the next section, the potential role of personality is discussed.

## 5.2 Person factors

How people behave in the social world is strongly influenced by their personality (Snyder & Ickes, 1985). Therefore, research started to investigate how individuals' dispositions influence their online social behavior. For instance, it was investigated how users' personality traits influence users' general social media usage (Correa et al., 2010), their self-presentation on social media (Seidman, 2013), or their online political engagement (Russo & Amnå, 2016). Recently, it was demonstrated that users' personality traits are better predictors for online political discussion behavior than their political ideology (Boulianne & Koc-Michalska, 2021).

Yet, these studies predominantly draw on one of the most prominent personality systems, the *Big Five* traits (i.e., openness to new experiences, extraversion, conscientiousness, agreeableness, neuroticism; McCrae & Costa, 2003). Indeed, these traits have proven to be comparatively stable over time (Rantanen et al., 2007; Soldz & Vaillant, 1999) and are also relatively robust concerning other person factors like sex or age (Soto,

2021). Nevertheless, other important personality systems can also have a crucial impact on individuals' behavior. In this vein, the *Dark Triad* personality traits gained noticeable attention in research when examining antisocial online behavior (for an overview, see Moor & Anderson, 2019). This personality model encompasses the subclinical levels of Machiavellianism, psychopathy, and narcissism found in all people to some degree (Paulhus & Williams, 2002). Machiavellianism reflects persons' cold, exploitative, and manipulative side of personality (Moor & Anderson, 2019; Paulhus & Williams, 2002). Narcissism reflects individuals' "normal" feelings of grandiosity, entitlement, dominance, and superiority (Paulhus & Williams, 2002, p. 557). Furthermore, psychopathy is characterized by a high propensity to impulsive behavior and thrill-seeking, whereas levels of empathy and anxiety are rather low (Paulhus & Williams, 2002).

For the online context, it was found that the dark triad correlates with diverse behaviors like compulsive Internet usage or self-presentation on social media (Petit & Carcioppolo, 2020). Moreover, it has been found that a range of antinormative online behaviors correlates with the dark triad. For example, all three traits correlate with cyberbullying (van Geel et al., 2017) and cyber-aggression (Pabian et al., 2015). Likewise, it was demonstrated that psychopathy and Machiavellianism are related to online trolling behavior (Lopes & Yu, 2017). Concerning incivility, results are mixed. Koban and colleagues (2018) did not find that any dark triad traits can predict uncivil online behavior. However, they only tested whether these traits predict concrete uncivil responses to provoking comments. When examining if the dark traits can predict whether people behave uncivilly at all, Frischlich and colleagues (2021) found that individuals' Machiavellianism and psychopathy are positive predictors for uncivil online behavior. Although narcissism did not affect uncivil participation in this study, it is still reasonable that narcissism plays a role in people's motives when commenting in an uncivil manner.

Thus, first studies indicate that the dark triad personality traits can foster incivility and related behavior. So far, little research has addressed how individuals' personality traits affect motivations for behaving in an uncivil manner. From a theoretical perspective, individuals' personality has a more substantial influence on their behavioral intentions than on their actual behavior (Ajzen, 1987, 1991).

### 5.3 First study

Taken together, the goals people pursue with uncivil online behavior are largely unexplored. Although some first insights exist on which motives people have when acting uncivilly online in general (Alonzo & Aiken, 2004; Moor et al., 2010), little is known about

individuals' motivation to behave in an uncivil way in online political discussions. Moreover, until now, no study has examined how the motives differ among distinct types of incivility. Therefore, it remains inconclusive to which degree people aim to communicate cooperatively in online political discussions when commenting in a specific uncivil manner. Consequently, the primary purpose of the first study is to explore the motives people have for distinct uncivil communication behavior in online political discussions and whether these motives are basically uncooperative or not.

However, as individuals' personality traits probably have a substantial impact on behavioral motives (Ajzen, 1987, 1991), it was also examined which role plays for acting in an uncivil manner. As the dark triad personality system was frequently associated with antinormative online behavior (Moor & Anderson, 2019), the focus lies on individuals' levels of psychopathy, Machiavellianism, and narcissism. Since research lacks insights into the motivations for uncivil discussion behavior, the study employs a qualitative research approach to explore which motives people identify for distinct uncivil. In addition, this approach is complemented with quantitative methods to provide more generalizable insights.

### ***5.3.1 Article 1: It's the aggression, stupid! An examination of commenters' motives for using incivility in online political discussions. (Kluck & Krämer, 2020)***

Drawing on the uses and gratification approach, the first study sought to identify motives people have when behaving in an uncivil manner in online political discussions. As incivility can take many forms, it was basically distinguished between different types of incivility as conceptualized by Bormann and colleagues (2021): (1) Disinformation and misleading exaggerations (i.e., violation of the information norm); (2) Insulting others, mocking others, or using vulgar speech (i.e., violation of relation norms); (3) Expressing antagonistic stereotypes, voicing threats against opposing groups, or denying opinion expressions of others (i.e., a violation against the context norm). Moreover, as less severe types of uncivil commenting behavior, (4) deviation from the discussion topic (i.e., a violation against the process norm), and (5) the usage of irony or sarcasm in comments (i.e., a violation against the modality norm) were included.

As a first step, it was of interest which motives people have when using one of the types of incivility in concrete situations. In this way, it was aimed to explore what people precisely drive to act uncivilly in online political discussions and which role situational aspects play. However, individuals have some overarching needs on a broader level that they want to gratify by participating in online discussions (Springer et al., 2015). These motives differ from specific commenting motives as they reflect individuals' more continuous

motivation to participate in online discussions. Therefore, these motives were assumed to be better predictors for explaining persistent (uncivil) behavioral patterns.

Although people's general discussion motives are more stable than their motives for sending specific comments, these can also change over time and be very situation-dependent. Hence, it is helpful to consider individuals' stable social needs. Here, Foulkes and colleagues (2014) argue that people experience certain social rewards when interacting with others. The authors reason that social rewards differ from social goals as they represent the hedonic value individuals experience when interacting with others and not which goals people want to achieve. Therefore, social rewards can be considered precursors of social goals that function on a higher level (Foulkes et al., 2014). By exploring different factors of social rewards, Foulkes and colleagues (2014) found that the hedonic value experienced from "being cruel, callous and using others for personal gains" is strongly related to the dark triad traits (p. 3). It was already demonstrated that this factor, which was labeled as *negative social potency*, can predict antinormative online behavior in the form of trolling (Craker & March, 2016). However, it has neither been investigated in which ways negative social potency affects lower-level behavioral motives nor has its role been investigated in uncivil online behavior. On the same level, peoples' personality predicts their behavioral intentions rather than the actual behavior (Ajzen, 1987, 1991). Therefore, personality traits have been considered to be more related to goal striving than to goal accomplishment (Parks & Guay, 2009). In the context of online communication, the dark triad personality traits seem to be drivers for antisocial online behavior (Moor & Anderson, 2019). As the dark triad's role is largely unexplored, the focus of the current study lies on the influence of psychopathy, Machiavellianism, and narcissism.

Taken together, motivation can be distinguished on different levels. People have certain goals they pursue at the micro-level when writing a specific comment or joining a concrete discussion. As little is known about individuals' particular goals when acting in an uncivil manner, the study's first aim was to explore individuals' motives for different manifestations of incivility. At the meso-level, people who actively discuss online have more general goals that continually lead them to online discussions. Hence, it was further of interest how these motives affect the frequency of uncivil discussion behavior. At the macro-level, in turn, the hedonic value people experience from specific social interactions or their levels of certain personality traits such as the dark triad reflect their inherent motivation to interact in the social world. Thus, it was expected that these variables affect uncivil commenting behavior through the pathway of antisocial discussion motives.

A survey-based study was conducted to investigate the research objectives addressing German-speaking individuals who actively discuss online ( $N = 115$ ). The study was divided into two central parts: First, using open-ended questions, peoples' motives for using the distinct types of incivility were explored within a qualitative approach. The resulting data were analyzed inductively so that higher-order categories were developed so that theoretical assumptions could be derived (Thomas, 2006). Second, within a quantitative approach, it was examined how participants' general discussion motives influence the frequency of uncivil commenting and how these motives are related to the dark triad and individuals' experienced value from negative social potency. These quantitative data were analyzed within a path model (AMOS 25) and a series of ordinary least square regressions (SPSS 25).

Although only people who reported actively discussing online were surveyed, all types of incivility were infrequently used by participants. However, participants reported seeing such behavior regularly. The analysis of the qualitative data revealed different categories of motives for uncivil discussion behavior: *Information related motives* (i.e., correcting information, adding information, or explaining certain facts), *discursive motives* (i.e., directing others' attention, exchanging with others, stimulating others to think), *identity-related motives* (i.e., signaling standpoints and group belongingness), *aggressive motives* (i.e., signaling intellectual inferiority, fueling conflicts, and annoying, provoking, mocking, embarrassing, or discrediting others), *counteractive motives* (i.e., objectification of emotional or irrational discussions, criticizing others' behavior or point of view, or paying back others improper behavior), and *affective motives* (i.e., expressing emotions, self-regulation as well as self-reflection, and having fun). Descriptive data showed that, for most types of incivility, people indicated having counteractive motives (i.e., insulting or mocking others, using antagonistic stereotypes, denying opinion expressions of others, deviating from the discussion topic, and using irony or sarcasm). However, for stating disinformation or misleading exaggerations, people predominantly indicated having information-related motives. Likewise, participants indicated mostly having affective discussion motives when commenting in a vulgar manner. For voicing threats against others, they stated most often having discursive or identity-related motives. Interestingly, aggressive motives were rarely indicated as a motive for uncivil behavior and did not predominate for any of the different types of incivility.

Overall, the qualitative data analysis implied that people predominantly aim to counteract antinormative or undesired behaviors of others when behaving in an uncivil manner. This tendency was already demonstrated in the context of flaming on YouTube (Moor et al., 2010). However, descriptive data of the current study suggest that this tendency

also applies to the context of online political discussions. Participants also reported having counteractive motivations for most types of incivility – not only for types of incivility related to flaming behavior. In this vein, Gervais (2017) reasoned that incivility is partly an anger-induced reaction to perceived uncivil attacks. Therefore, a broad range of uncivil behavior might be considered an appropriate way to penalize others' perceived competitive behavior.

Regarding the quantitative analyses, four distinct discussion motives have been identified by employing an explorative factor analysis: Cognitive discussion motives (i.e., knowledge gain), affective discussion motives (to compensate feelings of anxiety and sadness), discussion motives related to positioning (i.e., disagreeing, counteracting other comments, or expressing one's opinion), and aggressive discussion motives (i.e., being hostile towards others). A path model revealed that aggressive discussion motives were the strongest predictor for frequently behaving in an uncivil manner. However, motives of positioning also facilitated the usage of incivility. In contrast, affective motives decreased the frequency of uncivil discussion behavior, whereas cognitive motives did not influence the dependent variable. Moreover, the path model revealed that individuals' experience of negative social potency and individuals' levels of psychopathy and Machiavellianism positively predicted aggressive discussion motives and, therefore, indirectly increased uncivil commenting frequency. Consequently, additional regression analyses revealed that aggressive motives were the main driver for most types of incivility. Only for the usage of antagonistic stereotypes, having positioning-related motives was a slightly stronger predictor.

On the one hand, it appears contradictory to qualitative results that aggressive discussion motives had the most substantial effect on frequent uncivil online behavior to the findings of the qualitative analyses at first glance. On the other hand, positioning-related discussion motives also had a relatively strong positive effect on frequent uncivil discussion behavior – predominantly on antagonistic stereotyping. However, when examining the differential effects, only aggressive discussion motives predicted all types of incivility. In contrast, positioning-related motives only predicted mockery, the usage of antagonistic stereotyping, and being ironic or sarcastic. Thus, people seem to be more likely to derogate the “opposing side” when having positioning-related discussion motives.

### ***5.3.2 Conclusions from Study I and subsequent steps***

The first study of the thesis identified two predominant motives for acting uncivilly in online political discussions: Counteracting inappropriate discussion behavior was frequently indicated as commenting motivation when looking at single cases of uncivil behavior. From this perspective, incivility is rather driven by contextual factors as others' discussion behavior

provokes it. However, it was also found that general aggressive discussion motives predominantly drive frequent uncivil behavior. This pattern was found to be highly influenced by persons' "dark" dispositions. Therefore, routine uncivil behavior seems to be caused by person characteristics rather than contextual factors.

Regarding cooperative communication, it seems that those who frequently behave in an uncivil manner are not driven by cooperative goals but instead want to induce conflict in the discussion. On the other hand, as counteraction is a predominant motive for behaving uncivilly on a single case level, it can be assumed that people aim to punish violators of communicative norms and/or seek to enforce communicative norms that have been perceived to be violated (see also Tomasello, 2019, Chapter 9). Then, as people severely disapprove violations of communicative norm violations, it can be said that people react uncivilly to others' uncivil behavior (Bormann et al., 2021). This tendency seems to apply largely to all types of incivility, although minor variations were found. Please see Figure 1 for a schematic illustration of the findings.

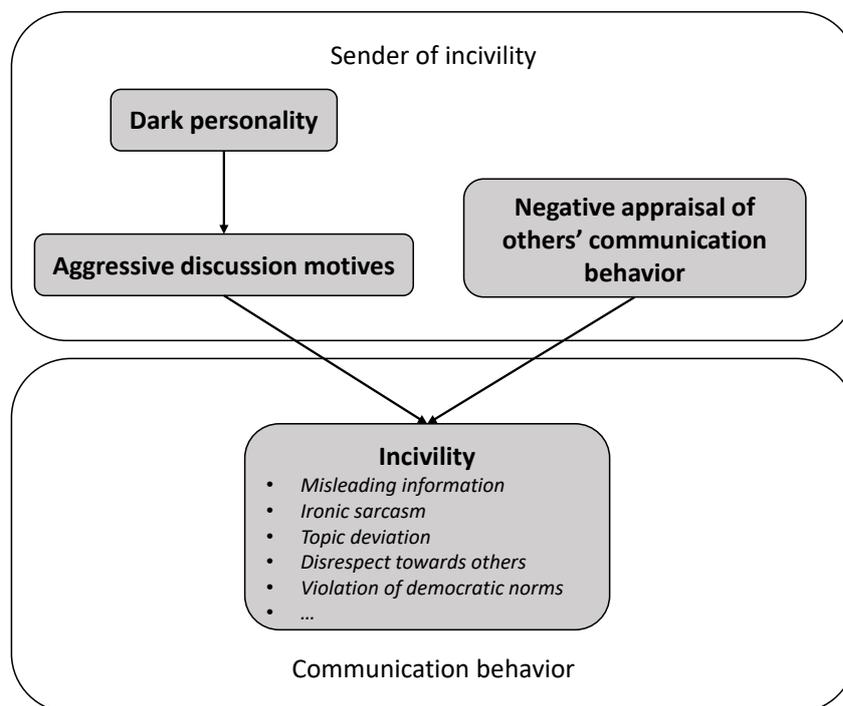


Figure 1. Schematic overview of the findings revealed by Study I. The dots indicate that the examined types of incivility are not exhaustive.

Thus, senders of incivility with rather malicious motives are predominantly driven by their “dark side” of personality. On the other side, individuals can also become the sender of an uncivil message when trying to counteract messages that they appraise as uncivil. The question arises of how they come to such a conclusion and why they regard incivility as proper means to react to perceived norm violations in discussions. Following Bormann and colleagues (2021), perceptions of incivility should result from recipients’ disappointed expectations of cooperative communication. Empirical validation for this assumption is, however, still missing. Moreover, as the first study of the current thesis focused on people’s motivation for discussing in an uncivil manner, it could not be examined how people react to different types of incivility. Hence, it remains unclear whether certain types of incivility trigger certain others or the same types of incivility. Based on these shortcomings, in the next step, the thesis examines how different types of incivility affect recipients’ appraisals of cooperation.

### **6 A lack of perceived cooperation? Appraising incivility in a social environment**

The first study gave valuable insights into individuals’ motives for behaving in an uncivil manner online. As no differences between the distinct types of incivility have been found, the sender’s motives are not helpful to categorize incivility by senders’ motives. However, as shown in Study I, distinct types of senders can be classified by their commenting motives. On the one hand, some commenters infrequently employ incivility to counteract perceived norm violations. On the other hand, some commenters frequently act in an uncivil way driven by their dark side of personality via aggressive discussion motives. That highlights that further research and interventions against incivility must be differently addressed based on the senders’ motives. Since users’ personality traits cannot be easily changed, it is more reasonable to find ways to cope with counteractive incivility in the first step. Although people who have counteractive motives seem to read uncivil content more often than they produce such content (see also Moor et al., 2010), uncivil behavior driven by counteractive motives is, in total, probably not seldom. This assumption is supported by findings that showed that preceding uncivil behavior and personal victimization is a stronger predictor of behaving uncivilly on a single case level than individuals’ dark personality traits (Frischlich et al., 2021). Likewise, Koban and colleagues (2018) found that single uncivil responses are rather driven by perceived provocation than by the dark triad personality traits. Therefore, on a single case level, provocations and preceding incivility seem to be a better predictor of uncivil online behavior than personality traits.

At first glance, this tendency might not appear very surprising. However, at the same time, this highlights the ambivalence of uncivil communication as people violate communication norms themselves to counteract others' violations of communicational norms. On a perceptual level, incivility is more than the disapproval of the usage of certain words. In the end, recipients' interpretation of the message determines whether they appraise a message as uncivil or not.

So far, little is known about the mechanisms that lead individuals to appraise a message as uncivil. As discussed in Chapter 2.2, the appraisal of incivility is usually measured in a unidimensional way and focuses on the message rather than on the sender of the message. However, the appraisal of the sender of an uncivil message is probably a better predictor of recipients' reaction than the mere perception of the uncivil message. Here, the basic approach of Bormann and colleagues (2021) is again helpful as the focus highlight the social nature of incivility.

Yet, although Bormann and colleagues (2021) highlight that humans' ability to share and read each others' intentions is a central prerequisite for cooperative (communication) behavior (see Tomasello et al., 2005 for a detailed discussion), the authors did not further address how uncivil online communication can affect this process. O'Sullivan and Flanagin (2003) reasoned that the perceived intention determines how people assess communication norm violating behavior. In the context of flaming, the authors argue that only corresponding norm violations that are perceived to be intended are considered as true flaming behavior. In contrast, unintentional norm violations would be more likely to be viewed by recipients as miscommunication. Thus, when defining incivility as uncooperative behavior that its recipients disapprove, the question arises of how people conclude that senders of norm violating messages are unwilling to cooperate and whether they react in another form when they think that the sender of a norm violating comment is basically willing to discuss. To explain how people conclude that a (potential) discussion partner is uncooperative, attribution theories are helpful. These theories have a long tradition in social psychology and basically suggest that people rely on certain attributions about others' intentions and/or dispositions to make sense out of their behavior. Drawing on these core assumptions, the next studies of the thesis addressed how people process incivility in online political discussions and how this influences their own behavior.

### **6.1 Making sense out of others' (uncivil) behavior – the role of attribution theories**

Heider (1958) was the first who noticed that attributions are a central process for forming subjective interpretations of the world and laid the foundation for subsequent

attribution theories, developed and refined in many ways. For a long time, attribution researchers assumed that people rely on the distinction between internal and external causes when explaining others' behavior (e.g., Kelley, 1967). By now, scholars emphasize the importance of perceived intentionality for behavior. Based on Heider's (1958) initial reasoning, it is argued that the internal-external dichotomy is only helpful to explain unintentional behavior or outcomes of actions (e.g., success or failure) because, in these cases, an actor lacks direct control. Intentional behavior, in contrast, is explained by the intentions and/or dispositions that are inferred to an actor by the receiver of specific behavior (Jones & Davis, 1965; Malle, 1999; Malle & Korman, 2013; Reeder, 2009; Reeder & Brewer, 1979; Weiner, 2000). That means, if people think that a person behaves intentionally, they do not evaluate whether the cause lies within internal (person) factors or external factors. Rather, people consider which goals the actor has and which reasons the person has for the action. Transferring this thought to online discussions, people will, of course, not assume that someone accidentally writes a comment. However, in terms of uncivil comments, they will probably question what motivates someone to violate norms of communication. According to the fundamental attribution error (Ross, 1977), people tend to overestimate the driving force of dispositional factors and underestimate the influence of situational cues.

Reeder (2009; see also Reeder et al., 2004) argues that people make multiple inferences when they observe others' intentional behavior. His proposed *multiple inference model* suggests that people search for others' intentions when they observe certain behavior. Also, the model suggests that people consider the situational factors under which the behavior is performed to identify these intentions accurately. From these inferences, people then make attributions about the persons' stable traits to obtain a more comprehensive picture of the causes of the behavior. Thus, these attributions help observers to understand others' behavior and, therefore, to assess a social situation (Reeder, 2009). Another attribution approach to describe how people explain others' behavior was introduced by Malle (1999). Following his *folk-conceptual theory of behavior explanation*, three modes of behavior explanation are used when people consider others' behavior as intentional: reason explanation, causal history of reason explanations, and explanations based on enabling factors. According to Malle (1999, 2011), the most common mode on which people rely when making sense of others' behavior is the mode of reason explanation. Then, people infer particular intentions from others' behavior, which is formed by the attributed beliefs and desires of the behaving person. However, Malle (1999, 2011) argues that sometimes people additionally rely on factors that lay behind the actors' reasons (i.e., causal history of reason explanation). This mode

addresses, for instance, attributions or knowledge about the actors' personality, cultural background, or unconscious mental states. According to Malle (1999, 2011), the third mode, enabling factors, does not directly explain why someone performed particular behavior but rather which factors enabled the person to perform the behavior (e.g., she trained a lot, so she could run the marathon). Of course, people do not rely exclusively on one mode but can also combine these modes to explain others' behavior.

Although the folk-conceptual theory of behavior explanation (Malle, 1999, 2011) and the multiple inference model (Reeder, 2009; Reeder et al., 2004) differ in certain ways, they are not mutually exclusive. Quite the contrary, the approaches can be combined to explain the appraisal of others' behavior more comprehensively. In this vein, a major shortcoming of both approaches is that the focus lies on the passive explanation of others' behavior, and the possible active role of the observer in a social situation is hardly considered. Consequently, how observers' attributions affect their own behavior is not further addressed. Primarily, Malle's (1999, 2011) approach focuses on the mere explanatory value of attributions. Therefore, he sharply distinguishes between dispositional attributions as addressed by the multiple inference model (i.e., making trait inferences from behavior) and causal attributions (i.e., why was the behavior performed; see Malle & Korman, 2013 for a detailed overview). Although he concedes that these approaches are related, he emphasized that trait inferences have little explanatory value. However, when observers decide how to react to others' behavior, both types of attributions are probably important. Especially in situations of cooperative interactions, people must make a range of attributions that determines their own behavior (Tomasello et al., 2005). The other way around, making the "wrong" attributions can lead to the perception of uncooperative behavior. Nevertheless, as a first step, Malle's (1999) approach seems to be more valuable in explaining how the uncivil behavior of others is evaluated in terms of cooperation because in such cases people usually have to make different attributions that offer explanatory value for successful interaction with the counterpart (Tomasello et al., 2005).

## 6.2 Second study

In summary, people must make diverse attributions to decide whether one or more interaction partners are cooperative communicators. As discussed at the beginning of this chapter, individuals must assess others' intentions to be cooperative and interpret others' behavior as ethical to achieve cooperative communication. Moreover, they must trust that the interaction partners have the skills to communicate cooperatively (Allwood et al., 2000). In terms of uncivil online behavior, the attributions which recipients of potential uncivil

messages make about the sender are likely to determine whether ideal cooperative communications emerge or whether cooperative communication fails at all. However, violations of communication norms can be very different (Bormann et al., 2021). Therefore, within the next study, it was examined how attributions differ between distinct types of incivility and how this affects recipients' own behavior. As it is unlikely that people make the same attributions for all types of incivility, we can gain knowledge of why people appraise specific behavior as uncivil and how this appraisal influences their discussion behavior. In this way, different types of incivility can be categorized by violations against norms of cooperative communication (Bormann et al., 2021) and, on a perceptual level, by which of recipients' expectations towards the sender are violated the most.

***6.2.1 Article 2: Unwilling, intolerable views, or incapable? – How different types of incivility can be distinguished based on attributions about cooperative communication (Kluck, Bormann, Rieß, & Krämer, 2021)***

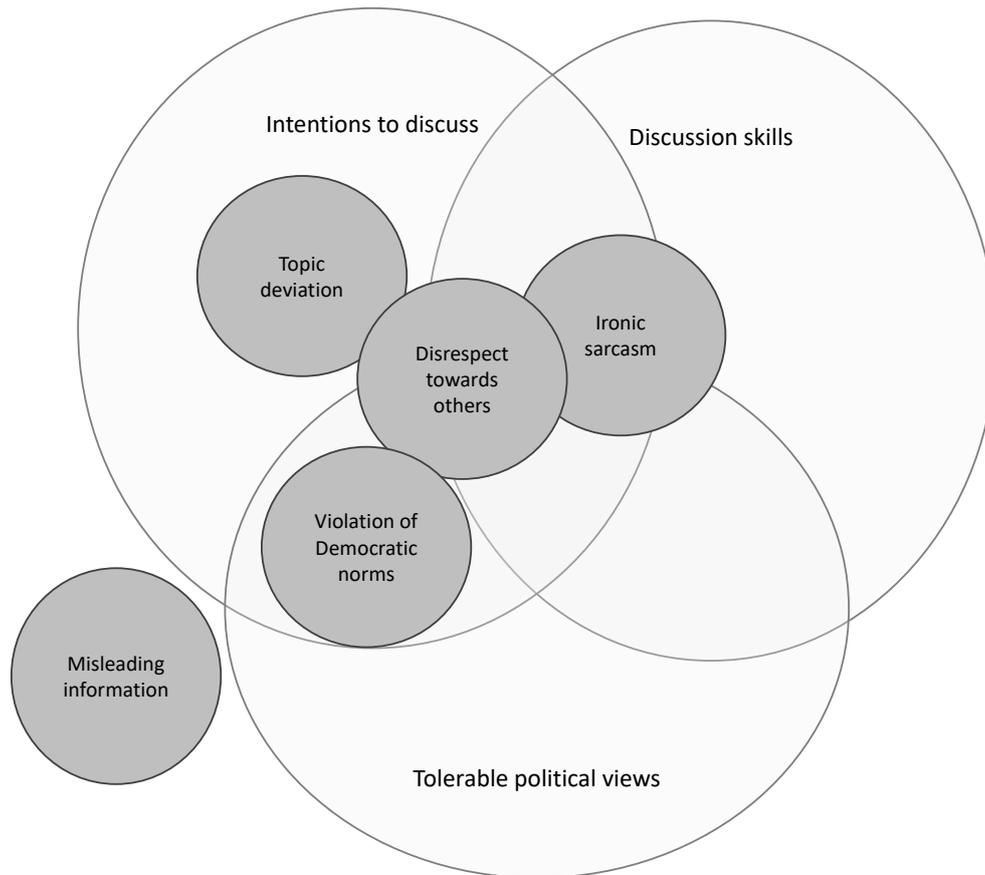
Within Study II, the assumptions about cooperation and cooperative communication (Allwood et al., 2000; Tomasello et al., 2005) and the folk-conceptual theory of behavior explanation (Malle, 1999, 2011) were combined. It was argued that in potential cooperative situations, people have to make certain inferences about others based on their behavior to decide whether they are suitable for cooperative interactions or not. More specifically, in the context of online discussions, it was argued that people need to appraise other commenters as cooperative discussion partners for a constructive discussion to occur. However, incivility was shown to decrease cooperative behavior, such as joining or staying in a discussion, respectfully acknowledging others' perspectives, and adding new perspectives to the discussion (Han et al., 2018; Han & Brazeal, 2015). Therefore, recipients of incivility are not likely to have trust that the sender will behave in a manner consistent with the requirements of cooperation as proposed by Allwood and colleagues (2000). Based on these assumptions and suggestions from Malle's (1999) folk-conceptual theory of behavior explanation, distinct dimensions of attributions have been derived in Study II. First, people must have a certain trust in others to have the intention to discuss. In line with the idea of ethical considerations from Allwood and colleagues (2000), they must also consider the perspectives of others to be tolerable. Finally, people must expect that discussion partners have appropriate discussion skills. Otherwise, cooperative communication is probably impeded, as described by Bormann and colleagues (2021). However, we expected that different types of incivility could lead to varying kinds of attributions. Moreover, we aimed to explore how the different types of incivility and connected attributions affect individuals' uncivil behavior, willingness to

answer, and constructive discussion behavior. In this way, we aimed to find new ways of categorizing incivility based on its outcomes.

An online experiment ( $N = 374$ ) was conducted to investigate the research objectives. Based on the different violations of cooperative norms as proposed by Bormann and colleagues (2021), it was tested how attributions differed between civil comments and comments, including misleading information, ironic sarcasm, topic deviation, disrespect towards other discussants, or violations of democratic norms. An allegedly real discussion forum – “Let’s Discuss!” – was specially constructed for the study. Specific discussion rules were developed to create a situation in which cooperative communication is required and ensure that participants answered the manipulated comment. The cover story told participants that the study was about testing the forum and that they would in fact discuss with other people.

As expected, results revealed that the distinct types of incivility lead to different attributions. An interesting finding was that disrespect towards others was consistently rated low on all attributional dimensions, whereas other types of incivility were inconsistently distributed across the different dimensions. For instance, the exposure to comments that violates democratic norms predominantly decreased attributions of tolerable political views and even did not decrease attributions of discussion skills compared to civil comments. Likewise, commenters who deviated from the discussion topic were attributed with fewer discussion intentions than civil comments but did not significantly differ regarding the other attributions. Thus, attributions, indeed, allow a new way of categorization. Seemingly, disrespect towards others is considered to be highly uncivil (Kenski et al., 2020; Muddiman, 2017; Stryker et al., 2016) as it fails to signal cooperation on various dimensions.

However, although seldomly conceptualized as uncivil, it was shown that topic deviation leads to decreased attributions of discussion intentions. As this attribution can probably impede further cooperative communication (Allwood et al., 2000; Tomasello et al., 2005), topic deviation can be deemed uncivil (Bormann et al., 2021), and it can be identified what the reason is. Likewise, as violations against democratic norms predominantly decreased attributions of tolerable views, cooperative communication might rather fail because people do not assess the person to communicate in an appropriately ethical manner. These patterns demonstrated that assessing distinct attributions about the sender allows a more nuanced way to measure and categorize incivility on a perceptual level. Please see Figure 2 for an illustrated summary.



*Figure 2.* Visual categorization of the different types of incivility based on recipients' attributions. The illustration is adopted from the submitted manuscript.

Regarding the behavioral outcomes, participants discussed the given issue primarily in an exemplary manner, and only a few uncivil commenting behavior cases occurred. Hence, we did not find that exposure to the distinct types of incivility appreciably affects people's usage of incivility. Only for topic deviation when being exposed to a topic deviation, people seemingly tend to further deviate from the discussion (i.e., introducing a point that neither addresses the content of the previous comment nor the topic of discussion). This outcome supports the assumption that topic deviation should not be underestimated as a type of incivility. However, none of the attributional variables affected this behavior. Therefore, this behavior can probably be "contagious" itself.

It also appeared that most participants decided to write an answer in the study's discussion forum. For those who chose to pass or to leave the discussion, no significant differences between the conditions could be found. Likewise, we did not find that people significantly used more or fewer arguments based on the comment they were exposed to.

However, results revealed that people tend to deviate from the discussion topic when they were exposed to a topic deviation before. We also found that people were especially more inclined to criticize individuals who disrespect other discussants or violated democratic norms. Additional analyses revealed that the less tolerable political view and discussion skills were attributed to the sender of the comment, the more participants were likely to express critique.

Interestingly, attributions about discussion intention did not affect critique expression. In this context, previous research demonstrated that people are more likely to punish intentional uncooperative behavior and rather teach individuals who are uncooperative due to ignorance (investigated on toddlers; Kachel et al., 2018). Therefore, on a speculative basis, this implies that critique expression triggered by uncivil commenting behavior might be more than mere punishment. Instead, people may aim to educate others when engaging in this behavior.

### ***6.2.2 Conclusions from Study II and next steps***

Study II demonstrated that uncivil online behavior is not appraised separately from its sender. Quite the contrary, inferences about a commenter can be very different depending on whether they comment in an uncivil way and which type of incivility they use. Therefore, the assumption that incivility can be categorized by recipients' attributions was supported. Especially disrespecting behavior was associated with low levels of cooperative behavior on all attributional levels. However, in contrast to the results of Study I, we could not support that the attributions triggered further uncivil behavior on a significant level. Only deviations from the discussion seemingly increase when people are exposed to the preceding topic deviations. Noteworthy, as indicated by Study I, uncivil commenting behavior is rarely performed by discussion participants (see also Moor et al., 2010) – even when counteracting incivility. Instead, we found that critique expressions can be triggered by decreased attributions of tolerable political views and discussion skills. Thus, people seem to be more likely to verbally disapprove of behavior that lies in peoples' person characteristics rather than their immediate intentions. See Figure 3 for an illustration of the findings.

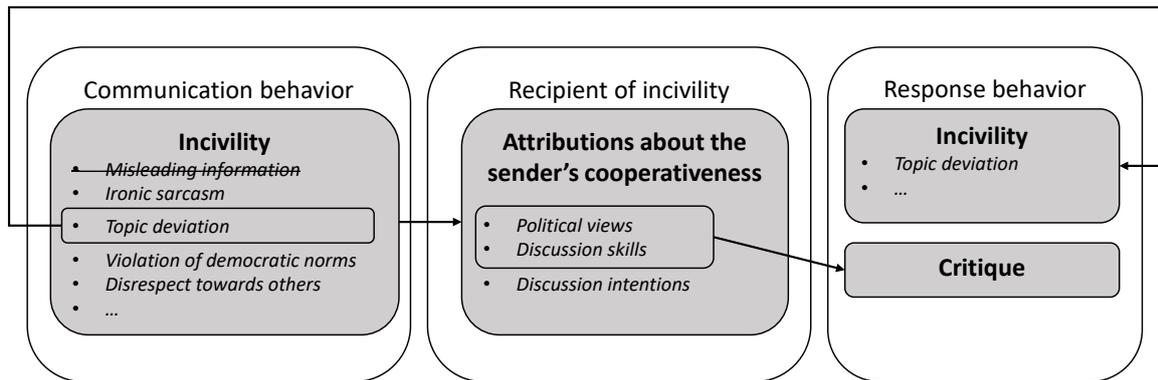


Figure 3. Visualization of the findings of Study II. Misleading information is crossed out, as it was not found that it influences attributions about cooperativeness. Moreover, it was not found that the usage of specific types of incivility is affected by the attribution process. However, recipients of topic deviation were slightly more likely to engage in further deviating behavior. In that cases, the norm violation itself probably has a direct contagious effect. The dots indicate that the examined types of incivility and findings are not exhaustive.

Nevertheless, we did not account for all possible motives which people can attribute to uncivil behavior. We explicitly investigated whether people attribute cooperative intentions (i.e., discussion intentions) to the sender of an uncivil message. Therefore, other attributed motives might also lead to critique expressions. Likewise, although we did not find that individuals were more likely to behave in an uncivil way when exposed to incivility, other studies have found such effects (Chen & Lu, 2017; Gervais, 2015; Kwon & Gruzd, 2017; Ziegele et al., 2018). Hence, it still can be assumed that the likelihood of critique expression can be increased by other attributions people make. As Study I revealed that an aggressive discussion intent increases the frequency of uncivil commenting behavior, the attribution of aggressive intent might cause further aggression (Gonçalves, 2018; Tomasello, 2019, Chapter 9).

However, it was shown that different types of incivility lead to distinct attribution patterns. In that sense, it is unlikely that people attribute an aggressive intent to individuals who deviate from the discussion topic. In Study II, topic deviation was predominantly associated with a lack of discussion intentions but did not affect attributions of tolerable views or discussion skills. Thus, such behavior seemingly does not lead to a comprehensive negative evaluation of the sender. Rather, people might focus on a lack of interest in the discussion topic. In contrast, incivility in the form of disrespect towards others was associated with the lowest attributions of discussion intent and significantly decreased attributions of tolerable political views and discussion intentions. As such behavior usually attacks individuals on a

personal level (Brown & Levinson, 1987; Chen & Lu, 2017), people might perceive more than just a lack of cooperative behavior. Indeed, in Study I, aggressive discussion motives had the strongest effects on disrespecting behavior in the form of insults and vulgar expressions. Based on these considerations, in the next step, a broader understanding of the motives that people attribute to others who engage in disrespecting behavior should be generated.

### **7 The role of attributed aggressive intent: Incivility within the general aggression model**

So far, the first two studies focused on the question of why people use distinct types of incivility and which attributions recipients make about the senders. However, people can make a range of attributions that assumably lead to different outcomes. Here, the intentions people attribute to others seem to be crucial for how individuals explain others' (uncivil) behavior (Malle, 1999; O'Sullivan & Flanagin, 2003; Reeder, 2009). So far, the intentions for uncivil discussion behavior have hardly been investigated. It was argued that although uncivil behavior can be basically cooperative if communicators cognitively consider each other, it probably impedes ideal cooperative communication as it decreases trust in communication partners' ethical considerations (see Allwood et al., 2000). Then, a constructive discussion outcome is rather unlikely. In other words, when behaving in an uncivil way, people can cooperatively attack each other but then will probably not cooperatively discuss an issue. Then, people must first conclude that their (potential) discussion partner has intentions like attacking others or disrupting the discussion. As the results of Study II suggest, such inferences might occur especially based on disrespectful behavior. How people can come to such a conclusion and how this process affects their own behavior is described well by the *general aggression model* (Anderson & Bushman, 2002).

The general aggression model (Anderson & Bushman, 2002) allows a more nuanced perspective on how individuals process uncivil discussion contributions. Combining theories addressing social interaction and cognitive processing, the model was initially developed to describe how aggression evolves in humans and leads to violent behavior. However, it is also valuable for explaining nonviolent behavior due to its integrative approach (Buckley & Anderson, 2006). Broadly outlined, the model (Anderson & Bushman, 2002) proposes a three stages process of how individuals are affected by certain (aversive) events. In the first stage, inputs are given by the situation and an individual's person characteristics. In turn, these inputs affect individuals' cognitions, emotions, and arousal in the second stage. Finally, in the third stage, people appraise the given situation and decide how to react to a related event. A core idea of the general aggression model (Anderson & Bushman, 2002) is that individuals develop specific knowledge structures based on their experiences, guiding how the different

steps proceed. Some knowledge structures are emphasized as particularly important: perceptual schemata (i.e., how to perceive entities and events), person schemata (i.e., beliefs and attitudes towards other individuals and social groups), and behavioral scripts (i.e., knowledge about which behavior is appropriate for a particular situation). However, the list can be continued. For instance, Allen et al. (2018) added an individual's beliefs and attitudes to the list of important knowledge structures. Thus, the model proposes that knowledge structures guide how contextual factors and person characteristics influence an individual's cognitions, emotions, and arousal, which in turn influence the person's appraisal of the situation and the related reaction. Importantly, Anderson and Bushman (2002) do not propose that this process follow a strict order. Instead, the model suggests that the internal state variables (i.e., cognitions, emotions, and arousal) and the appraisal of the situation can influence each other mutually. This model may also prove helpful for understanding in which situations incivility has a destructive effect on discussion dynamics – such as discouraging others from participating or resulting in reciprocal uncivil attacks. Indeed, many findings related to incivility can be integrated into the general aggression model (Anderson & Bushman, 2002).

Concerning contextual factors, it was found that whether distinct types of incivility (i.e., insults or mockery) increase negative emotions depends on the topic being discussed (Wang & Silva, 2018). Likewise, the media platform on which incivility occurs can shape its awareness (Sydnor, 2018). Studies also demonstrated that the recipient's person characteristics are essential for the processing of incivility. Sydnor (2019) found that conflict-avoidant individuals are more likely to suffer negative consequences from incivility in contrast to conflict-oriented individuals. Furthermore, Kenski and colleagues (2017) found that people high in agreeableness are more likely to assess incivility as uncivil, whereas conservative individuals were less likely to consider incivility as actually uncivil.

Further supporting the assumptions of the general aggression model (Anderson & Bushman, 2002), there is ample evidence that incivility can influence individuals' emotional states (Masullo et al., 2021; Gervais, 2015, 2017; Hutchens et al., 2019; Lu & Myrick, 2016), and cognitions (Rösner et al., 2016). By contrast, little attention has been paid to how its recipients eventually appraise incivility. Though, this stage is crucial because, at this point, it is decided how recipients will react to an uncivil message. Here, Anderson and Bushman (2002) differentiate between immediate appraisals and reappraisals. The authors argue that immediate appraisals occur rather automatically, effortlessly, spontaneously, and unconsciously. In contrast, when individuals have sufficient resources and regard their

reaction to be important, the authors argue that they are likely to make a reappraisal. Hence, the reappraisal proceeds more conscious and effortful. Anderson and Bushman (2002) state that individuals then search for alternative interpretations of the situation and choose their reactions more carefully. Nevertheless, how individuals come to the respective appraisals is only described broadly within the general aggression model (Anderson & Bushman, 2002). Allen and colleagues (2018) explain that

immediate appraisals often include trait or situational inferences. For example, if a man bumps into a woman at a crowded party, she could make a trait inference (e.g., “He meant to do that— what a jerk!”) or a situational inference (e.g., “It’s so crowded—I’m sure that was an accident.”). (p. 77)

Although this complies with the assumption that attribution theories are helpful to explain how people appraise a social event (see Chapter 6.1), this is a very abbreviated description of a more complex attribution process. To be more precise, the notion “He meant to do that— what a jerk!” includes related but different attributions: The recipient makes attributions about the man's intentions for bumping into a woman and then makes conclusions about his personality. That is the process described by the multiple inference model (Reeder, 2009). As outlined in Chapter 6.1, the model suggests that people search for the behavioral motives of other agents, considering situational factors. Within the context of the general aggression model (Anderson & Bushman, 2002), this approach indeed seems to be more valuable than Malle's (1999) folk-conceptual theory of behavior explanation, since here the agents' behavioral motives are decisive in the first place. Although the multiple inference model (Reeder, 2009) does not further address how the distinct attributions influence observers' own behavior, the general aggression model (Anderson & Bushman, 2002) indicates that the assessment can be crucially for individuals' reactions, respectively. When people come to the appraisal that someone is a “jerk” or an “idiot,” based on observed behavior, uncooperative responses are assumably more likely than cooperative ones. Even more, in accordance with the general aggression model (Anderson & Bushman, 2002), such appraisals likely encourage individuals to react aggressively. However, it has not yet been investigated how uncivil commenting behavior facilitates such an assessment and which motive attributions promote this.

In this vein, people must first perceive a comment to be uncivil. This is not as trivial as it may sound. As incivility is a perceptual construct, people must consider that a communicative norm is violated (Bormann et al., 2021). That said, following the general

aggression model (Anderson & Bushman, 2002), individuals' knowledge structures are crucial for how people come to such a perception. These structures include distinct schemata that allow individuals to classify certain events. Therefore, a closer look at which factors of online discussion might affect the perception of incivility is necessary. Two factors seem to be especially important for the assessment of an uncivil comment: The represented stance and the appearance of other comments. In the following, it is discussed why these factors are important and in which ways they can influence the perception of incivility.

### 7.1 The role of others' stance

One significant advantage of online discourses is seen in enhancing opinion heterogeneity in online political discussions (Brundidge, 2010). Indeed, people are likely to be confronted with opinions that differ from their own stance within online discussions (Barnidge, 2015; Kim et al., 2013), which is also recognized by users (Barnidge, 2017). In this sense, Mouffe (2000, Chapter 4) makes a sharp distinction between agonism and antagonism (see also Bormann et al., 2021). She argues that “antagonism is struggle between enemies while agonism is struggle between adversaries” (p. 102-103). Hence, having different opinions does not necessarily delegitimize the other side. Only when individuals aim to destroy the other side, the democratic exchange is impeded (Mouffe, 2000). This view is consistent with the notion in Chapter 4 that cooperative discussions need not be like-minded. Quite the contrary, pluralism and diversity are seen as essential parts of the democratic exchange between citizens (Brundidge, 2010). However, this idea follows a theoretical ideal. Although disagreement, indeed, can have positive outcomes for online discussions, it also can bias individuals' perceptions of others and can discourage them to participate in discussions (Chen, 2017, see also Chapter 2). Whether disagreement in online political discussions is perceived in an antagonistic or agonistic way is arguably a fine line. Here, incivility might make a crucial difference. Foremost, incivility in the form of disrespecting behavior might signal others that they aim to see others as delegitimize enemies instead of adversaries as stated by Mouffe (2000).

So far, it is pretty unclear which influence the represented stance within an uncivil comment has on its perception and outcomes. As one of the first, Gervais (2015) highlighted this aspect. He found that opposing incivility is more likely to increase anger and aversion, whereas like-minded incivility is more likely to trigger further incivility. However, in a subsequent study, Gervais (2017) showed that opposing incivility triggers further uncivil behavior, whereas like-minded incivility does not lead to more incivility. In another study, Chen and Lu (2017) found that disagreement has a “chilling effect” on discussions by

increasing negative affect and aggressive intentions (p. 108). Hutchens and colleagues (2019) demonstrated that compared to an uncivil discussion, people's enthusiasm is increased by a civil conversation only if the comment comes from a group they do not belong to (liberals vs. conservatives). Likewise, Muddiman (2017) found that the same uncivil behavior is assessed as more uncivil when it comes from individuals of an opposing political party. Therefore, although it is not clear in which ways, it can be noted that the represented stance of an uncivil comment influences its processing and its consequences.

To better understand how an uncivil comment's stance affects recipients, basic social psychological theories addressing intergroup conflict combined with assumptions from attribution theory might be vital. According to the *self-categorization theory* by Turner and colleagues (1987), people favor individuals from an ingroup when a specific social identity is made salient. As within online political discussions group affiliations and opinions towards particular topics are often close together (e.g., Kim, 2015), it is likely that disagreement in online discussions causes certain feelings of group belongingness. Then, attributions about others are biased negatively. Kenworthy and Miller (2002) showed that individuals explain the attitudes of individuals from an outgroup with more external, more emotional, and fewer rational reasons.

Furthermore, in the context of the Iraq war, Reeder and colleagues (2005) demonstrated that people who were against the war attributed more selfish motives and less morality to people who favored the war. Therefore, it can be assumed that participants' stances in online political discussions attribution processes are biased. When people are exposed to opposing views, feelings of cognitive dissonance (Festinger, 1957) are likely to arise as recipients see their stance challenged (see also Chen, 2017, Chapter 2). To relieve this feeling, people may derogate individuals who oppose their views (Cooper & Mackie, 1983). Indeed, in Study I, it was found that people who aim to position themselves in a discussion are more likely to use types of incivility that disparage others (e.g., mockery). Hence, making negative attributions about others might help individuals cope with the cognitive dissonance induced by opposing opinions.

So far, it has not been investigated whether such effects can alter the attributions made to uncivil commenters. That might be an essential step to understanding the inconsistent findings regarding the impact of disagreement. In particular, capturing the motives people attribute to others who disagree with them can be crucial. Then, it can be examined whether incivility in combination with disagreement facilitates the perceptions of antagonism as proposed by Mouffe (2000).

## 7.2 The role of discussion norms

Another factor that is probably important for an individual's appraisal of uncivil messages lies in the salient discussion norms. In online political discussions, guidelines like "netiquettes" usually prompt users to use a respectful tone and respect democratic norms. Moreover, as discussed in Chapter 2.1, people typically have certain normative expectations regarding how communication should proceed in (online) political discussions (Bormann et al., 2021). However, this is not necessarily the prevailing norm in all discussions. Chen and colleagues (2019) note that individuals' uncivil commenting behavior could sometimes be regarded as appropriate. Therefore, in Study II, cooperative norms have been highlighted accordingly to ensure that the distinct types of incivility are perceived as corresponding norm violations of cooperative communication.

Indeed, as incivility can trigger further incivility (see also Chen & Lu, 2017; Gervais, 2015; Ziegele et al., 2018), it was concluded that offensive comments could shape the communicative norms within a discussion (Kwon & Gruzd, 2017). Shmargad and colleagues (2021) recently found that people are more likely to engage in repeated uncivil behavior when preceding comments were also uncivil and when others approved their uncivil commenting with upvotes. Another study by Rösner and colleagues (2016) suggests that an increasing number of uncivil comments mitigates the adverse effects of incivility. The authors demonstrated that exposure to an uncivil discussion increases individuals' hostile cognitions compared to a civil discussion. However, against their predictions, such cognitions did not increase with the number of uncivil comments in a discussion. Means even declined when the discussion consisted only of uncivil comments. Although the effect was not significant, this finding corresponds to the suggestion that incivility can become normative when it occurs more frequently (Kwon & Gruzd, 2017; Shmargad et al., 2021). This also corresponds to the conceptualization of incivility provided by Bormann and colleagues (2021) which highlights incivility as a perceptual construct. Therefore, if recipients of supposed uncivil behavior do not associate a norm violation with this behavior, it is, in principle, not uncivil. In other words, if uncivil comments define the communicational norm, they are not uncivil anymore on a perceptual level. Rather, overly polite comments are then likely to be viewed as violating cooperative communication norms.

Noteworthy, this idea is not new. Following the *focus theory of normative conduct* (Cialdini et al., 1990), social norms can be distinguished between injunctive norms and descriptive norms. Injunctive norms describe what people think others typically approve or disapprove of. Descriptive norms, on the other hand, refer to the actual behavior that is

observed in others. In terms of incivility, injunctive norms usually disapprove of such behavior as it violates expectations about cooperative communication (e.g., Bormann et al., 2021) or violates the policies typically set by discussion boards' (e.g., netiquettes). In contrast, when most or all participants of a discussion communicate in a supposedly uncivil manner, the descriptive norm would be to behave in such a way. It is important to note that these two norms are not mutually exclusive. However, which of the norms has the more substantial impact crucially depends on which norm is perceived to be more salient (Cialdini et al., 1990). Anyway, the aforementioned empirical findings (Kwon & Gruzd, 2017; Rösner et al., 2016; Shmargad et al., 2021) indicate that uncivil commenting as descriptive norms can override injunctive norms of polite and respectful communication. This assumption, again, highlights that incivility is more than a message characteristic. Nevertheless, more research is needed to understand whether people regard frequent uncivil behavior to be acceptable or incivility is contagious due to hostility attribution.

### 7.3 Third and fourth study

In sum, the general aggression model (Anderson & Bushman, 2002) provides a useful framework to investigate further how individuals process incivility. Consistent with the previous argumentation, the model highlights the role of attributions for individuals' reactions to others' (antinormative) behavior. Since in Study II the attribution of aggressive intentions was neglected, the role of such an appraisal should be investigated in the following steps. Although it was found that especially disrespect towards others leads to low perceptions of cooperation, this did not affect participants' discussion behavior. Therefore, it should be investigated more comprehensively which commenting motives people attribute to senders of such messages. Since the attributions of motives can lead to a more general evaluation of the sender (Reeder, 2009), it was of particular interest which attributed motives were most likely to result in a negative appraisal of the sender: the attribution of aggressive motives or a lack of attribution of discussion-centered motives. This should help understand why in Study II, disrespecting behavior had been attributed to low levels of cooperation on a rather universal level.

As explained in Chapter 7.1, the stance represented in an uncivil comment might bias the attribution process as well. Here, it might make a difference whether people disrespect the "own side" or "the others." Therefore, a further purpose of the study was to investigate which influence an (uncivil) comment's stance has on the attribution process.

Moreover, the general aggression model (Anderson & Bushman, 2002) suggests that the resulting attributions affect individuals' reactions. Hence, it was also of interest how

individuals' discussion behavior is affected by negative appraisals of the sender of an uncivil message. Although we did not find such effects within Study II, it was still assumable that uncivil messages – especially in the form of disrespect – can have such consequences. As a significant limitation of the study was that the discussion platform was fictional and had specific discussion rules, it was assumed that a more familiar setting would expand the findings of the attribution process and how this affects individuals' willingness to participate in discussions.

### ***7.3.1 Article 3: “What an idiot!” – How the appraisal of the writer of an uncivil comment impacts discussion behavior (Kluck & Krämer, 2021)***

Drawing on the assumptions of the general aggression model (Anderson & Bushman, 2002) and the multiple inference model (Reeder, 2009; Reeder et al., 2002), it was expected that people attribute more aggressive discussion motives to an uncivil commenter than nonaggressive (i.e., discussion-centered motives). Moreover, it was assumed that individuals then make more negative trait inferences through the pathway of the attributed discussion motives. In other words, it was examined whether people conclude that someone is an “idiot” through the attribution of distinct discussion motives. Based on the assumptions discussed in Chapter 7.1, it was further suggested that the stance of an uncivil comment could be a critical influencing variable of the attribution process. Following the self-categorization theory (Turner et al., 1987) and assumptions about cognitive dissonance (Festinger, 1957), we expected further that people attribute more aggressive and less nonaggressive discussion motives to commenters with an opposing stance than to commenters with a like-minded stance. Again, we additionally expected that, through this effect, inferences about negative traits increase. Since few consistent findings exist on the interaction of incivility and disagreement, it should be explored how these factors together affect the attribution process. Following the general aggression model (Anderson & Bushman, 2002) and previous empirical findings (Han & Brazeal, 2015), it was further assumed that as a result of the attribution process, hostile trait inferences discourage individuals from taking part in the discussion. Moreover, based on the behavioral outcomes of incivility that had been found earlier (Chen & Lu, 2017; Gervais, 2015; Ziegele et al., 2018) it was suggested that these inferences increase the likelihood that participants write an uncivil comment in the form of disrespect and that they express critique about the preceding incivility.

Finally, the general aggression model (Anderson & Bushman, 2002) highlights the role of individuals' characteristics, and Study I found that the Dark Triad personality traits (see Chapter 5.2) can positively affect the usage of incivility. Therefore, it was also of interest how

these traits function on a perceptual level. The literature indicates that psychopathy and Machiavellianism can be related to a lack of empathy (Ali & Chamorro-Premuzic, 2010). Likewise, it was found that psychopathy negatively affects the ability to identify interpersonal threats. Consequently, it was assumed that both traits decrease the likelihood that individuals attribute aggressive motives to senders of an uncivil message. For narcissism, such an effect could not be derived. However, narcissism was expected to influence attributions based on the comment's stance: As people high in narcissism are more vulnerable to ego threats (Li et al., 2016), and a dissonant opinion can lead to a related aversive cognitive state (Festinger, 1957), it was expected that people high in narcissism are more likely to attribute aggressive discussion motives to people opposing their stance.

An online experiment with 452 participants was conducted to investigate the research objectives with a 2 (civil vs. uncivil) x 2 (like-minded vs. opposing stance) between-subjects design. This time, participants were exposed to a mock-up discussion on Twitter. Individuals saw a Tweet asking about others' opinions. Under the Tweet, a single answer was manipulated according to the conditions. Two distinct topics were employed to increase the generalizability but were not treated as an additional factor.

As disrespecting behavior can manifest in different ways, the comments in the uncivil condition included either name-calling or vulgarity. As there was no difference in the perception of the two forms of disrespecting behavior, they were merged into a single factor in the analyses. To measure the attributed discussion motives, scales were constructed in an explorative way as a first step. By conducting an exploratory factor analysis and a confirmatory factor analysis, three different discussion motives could be extracted: The first factor directly addressed aggressive motives, whereas the latter factors, interest-related and counteractive motives, were employed as nonaggressive discussion motives as both can be considered as norm-conforming.

The results of the analyses largely supported the theoretical assumptions. Both uncivil comments and opposing stances positively influenced hostile trait inferences. It was also shown that the exposure to uncivil comments significantly and strongly increased attributions of aggressive discussion motives and decreased the attribution of nonaggressive comments. Likewise, it was found that the exposure to an opposing stance positively affected the attribution of aggressive motives, albeit much more weakly than uncivil comments. Regarding the nonaggressive motives, being confronted with an opposing stance only decreased interest-related motives but not counteractive motives. Hostile trait inferences were, however, only affected by aggressive motives.

Nevertheless, the indirect effects of both conditions on hostile trait inferences through this path have been supported. Interestingly, the interaction of the conditions yielded no significant impact. These results again supported the assumption that people do not simply assess an uncivil message but also assess its sender. The effect sizes were remarkably large. Interestingly, the attribution of aggressive motives and interest-related motives (which came closest to cooperative intentions, as it addressed interest in other views) were about the same strength. However, as only the attribution of aggressive motives affected hostile trait inferences and, in turn, these inferences influence individuals' commenting behavior, it is reasonable to assume that aggressive attributions and a more general negative appraisal of people can better explain outcomes of incivility than the lack of perception of cooperation. In addition, the study provides evidence that incivility and disagreement yield similar effects but function independently. This finding is quite important as it is the first study investigating this issue within assumptions of social perception. In this vein, both incivility and disagreement seemingly cause a negative state that increases negative appraisals about its sender.

Regarding the behavioral outcomes, participants' willingness to participate in the discussion was indeed decreased, albeit the effect was small. Supporting the findings of Study II, hostile trait inferences only increased the likelihood of critique expression but did not increase uncivil commenting behavior in the form of disrespecting behavior. Here, Study III demonstrated that the attribution could indeed explain in parts the negative consequences of incivility. Therefore, the study provides a valuable base on which future interventions can build when designing interventions against incivility. First, interventions can try to alter attribution about others to mitigate such adverse outcomes. Second, as discussed earlier, the attribution theory approach can help identify whether uncivil comments are perceived to be offensive beyond their mere appearance. Finally, moderation analyses revealed that people who are very high in psychopathy did not attribute more aggressive intentions to senders of an uncivil message than those who wrote a civil one. However, the moderation effect was supported for neither Machiavellianism nor narcissism. Thus, in contrast to individuals' uncivil commenting behavior (see Study I), the dark traits have comparatively small to no effects on the perceptual level.

### ***7.3.2 Conclusions from Study III and purposes of Study IV***

In sum, Study III indicated that the attribution of aggressive motives could play an important role in the consequences of incivility. As only aggressive intentions led to an overall negative evaluation of the appraised commenter, these attributions seem to be more important than a perceived lack of cooperative intent as discussed earlier (see Chapter 4).

However, attributing an aggressive discussion intent does not exclude that the lack of perceived cooperative purpose alone can have a negative impact. For instance, Study II found that deviating from the discussion topic can encourage individuals to further deviate from the topic. It was also shown that the violation of democratic norms tends to trigger more of such norm violations. Since Study III focused on incivility in the form of disrespectful discussion behavior, the results must be viewed primarily in this context.

Nevertheless, this study demonstrated how important motive attributions are for the appraisal of an uncivil commenter. Regarding the categorization of incivility based on different attributions as investigated in Study II, Study III provides additional insight into how attribution processes might alter based on the type of incivility. Thus, the perspective of cooperative communication as introduced by Bormann and colleagues (2021) provides an integrative framework useful for comparing the consequences of different types of incivility. However, as shown in this study, for incivility in the form of disrespecting behavior, the attributed aggressiveness seems to be more critical than the lack of cooperative intent. Of course, this is not mutually exclusive but it shows that types of incivility are probably still evaluated on a different basis at a micro-level than at the cooperative level. Therefore, categorizing different types of incivility should be extended to whether people attribute aggressive intentions to the sender since there are probably essential differences between the distinct types.

Study III also broadens the understanding of the attribution process. Independently from incivility, an opposing stance negatively biased the motive attributions made by participants. From a theoretical perspective, it is conceivable that both manipulations increased perceptions of antagonism instead of agonism, as explained by Mouffe (2000). Then, the study results can be interpreted in the way that incivility in the form of disrespect signals that others are seen as enemies and not as legitimate adversaries – regardless of which opinion is represented. As an opposing stance provoked similar, albeit much weaker, effects, it might be that incivility alone causes a stronger perception of antagonism than an opposed opinion. Noteworthy, as such effects were not directly examined, these considerations are on a theoretical basis. Nevertheless, these findings are significant as they further shed light on the relationship between incivility and disagreement.

Finally, it was found that the Dark Triad personality traits have little impact on a perceptual level. This is also an interesting finding, as in Study I, psychopathy and Machiavellianism had a strong effect on uncivil commenting behavior. Therefore, the results support the assumption that two ways of the causes of incivility can be separated: Incivility

driven by individuals' dark person characteristics and incivility caused by preceding incivility.

Overall, Study III provides important insights into how incivility in the form of disrespect can affect the attributions about others in online discussions. Here, the attribution of aggressive motives seemingly plays a key role. However, some questions remained unanswered: As discussed in Chapter 7.2, the salient discussion norm is likely to affect which attributions people make about an uncivil commenter. People might attribute fewer aggressive discussion motives to individuals who join an uncivil discussion in an uncivil way compared to individuals who join a civil discussion in an uncivil way. Therefore, in Study IV, it should be investigated how the extent of incivility in preceding comments affects the appraisal of an uncivil commenter. Another shortcoming of Study III was that it did not investigate how emotions and cognitions impact the attribution of discussion motives. However, following the general aggression model (Anderson & Bushman, 2002), this pathway is crucial for individuals' appraisal of a particular event. This aspect was assumed to be especially crucial to understanding the effects of preceding (un)civil comments. Finally, although it was found that hostile trait inferences can decrease recipients' willingness to join the discussion and increases the likelihood of critique expressions, uncivil answers were not more likely. Hence, something more is needed that people engage in such behavior as well. Thus, it should also be examined what eventually causes uncivil responses.

***7.3.3 Article 4: Appraising Uncivil Comments in Online Political Discussions: How do preceding incivility and sender's stance affect the processing of an uncivil comment? (Kluck & Krämer, 2020)***

Based on the assumptions of the focus theory of normative conduct (Cialdini et al., 1990), the main purpose of the study was to investigate whether incivility in preceding comments and the appraisal of an uncivil comment. In line with the general aggression model (Anderson & Bushman, 2002) and literature on emotions in the political context (Marcus et al., 2000, 2006), research on incivility has identified distinct emotions affected by uncivil discussion comments. This includes increased negative emotions like anger (Masullo et al., 2021; Gervais, 2015, 2017) and anxiety (Lu & Myrick, 2016), as well as reduced feelings of enthusiasm (Hutchens et al., 2019). Likewise, as discussed in Chapter 7.2, Rösner and colleagues (2016) found that incivility can increase recipients' hostile cognitions. Therefore, it was assumed that people have fewer feelings of anger, anxiety, higher levels of enthusiasm, and fewer hostile cognitions when exposed to an uncivil comment accompanied by other uncivil comments compared to civil ones. Likewise, people were expected to then attribute

less aggressive motives to the sender of an uncivil message and that individuals' emotions and hostile cognitions mediate this effect. While Study III did not find that attributions based on uncivil comment behavior depended on the stance of the commenter, like-minded individuals might have a more substantial impact on the perception of norms than people with an opposing stance (Hogg & Reid, 2006). Consequently, it was of further interest whether the represented stance would influence the expected effects of preceding comments. Finally, in line with Study III, it was investigated how the attribution of aggressive discussion motives affects expressions of critique and uncivil commenting behavior in the form of disrespecting behavior. However, additionally, Study IV asked for participants' discussion intentions expecting that the attribution of aggressive discussion motives increases their aggressive answering intention and decreases discussion-centered answering intentions. As it is well established that intentions can predict actual behavior (Ajzen, 1991), it was also investigated whether the expected influence of motive attributions on the uncivil commenting behavior and critique expression is mediated by the answering intentions.

An online experiment ( $N = 611$ ) with a 3 (uncivil vs. civil vs. no preceding comments) x 2 (opposing vs. conforming recipient's views) between-subjects design was conducted to investigate our assumptions and research questions. Within an online survey, participants were presented with a mock Facebook post asking them what they thought about a specific topic that was socially relevant. Like in Study III, participants had been exposed to one of two topics to increase the generalizability of the findings. In line with the experimental design, participants were exposed to an uncivil comment that should be appraised later on. The comment was either preceded by five uncivil comments, five civil comments, or, as control, by five blurred comments. In addition, all of the comments either agreed or disagreed with the participants' stance on the discussion topic.

To investigate the assumptions, a path model was conducted. However, the data provided a poor fit to the proposed theoretical model. Hence, an alternative model has been tested, which was derived from assumptions of the general aggression model (Anderson & Bushman, 2002). The model suggests that individuals can engage in a reappraisal that, in turn, can alter initial emotions and cognitions when people when enough motivation and resources are present. As emotions and cognitions were not measured immediately, it was deemed conceivable that such a reappraisal may have occurred. Therefore, in a well-fitting alternative model, it was tested how participants' attributions about the senders' motives affect their emotions and cognitions and how these variables affect participants' discussion intentions.

Replicating findings of Study III, results revealed that people who were confronted with an opposing stance were more likely to attribute aggressive motives to the sender of the last uncivil comment (i.e., the comment that was appraised by participants). It was revealed that for attributions of aggressive discussion motives, it did not matter whether preceding comments were uncivil, too. This suggests that people appraise incivility in the form of disrespect independently of others' uncivil behavior. On the one hand, the perceptions of descriptive norms (how do others behave?) as proposed by Cialdini and colleagues (1990) might not have been strong enough. On the other hand, it is also conceivable that the injunctive norm (what is approved?) is more firmly rooted in people's knowledge structures and cannot be easily overridden by descriptive norms. Likewise, no interaction with comments' stance emerged, indicating that the represented stance in a comment plays an essential role in attributions of the sender's discussion motives but does not affect normative behavior.

In line with the assumptions of the general aggression model (Anderson & Bushman, 2002), it was also shown that attributions about the last commenter's aggressive discussion motives were positively related to participants' emotions (i.e., anger, anxiety, enthusiasm) and hostile cognitions. That enthusiasm also increased was surprising. As enthusiasm is related to positive goal pursuit (Marcus et al., 2000), it was assumed that people regard their general discussion behavior to be more purposeful, which might have triggered feelings of enthusiasm. Moreover, it was found that the attribution of aggressive discussion motives positively affects individuals' aggressive answering intentions. In addition, an indirect effect was suggested mainly through the pathway of individuals' feelings of anger. This finding supports that the general aggression model (Anderson & Bushman, 2002) is suitable to investigate the processing of uncivil messages. Surprisingly, however, in a similar strength, both anger and enthusiasm increased participants' intentions to answer in a discussion-centered way.

Consequently, an indirect positive effect of attributions about senders' aggressive discussion motives on discussion-centered answering intentions was found. Astonishingly, anger and enthusiasm increased these intentions to a similar extent. This can partly explain why incivility can lead to destructive and constructive behavior (Borah, 2014; Chen & Lu, 2017; see also Chapter 1 of the current thesis). When anger leads to more constructive behavior, people might try to neutralize this negative affective state (Marcus et al., 2006). From the perspective of the general aggression model (Anderson & Bushman, 2002), people then probably react in a more controlled and conscious way. Finally, it was found that

aggressive motive attributions positively predicted expressions of critique but not uncivil answering behavior. However, aggressive motive attributions indirectly increased uncivil answering behavior through the mediating way of aggressive answer intentions. Thus, whether incivility leads to aggressive answering intentions or not can make the difference. When people have high aggressive answering intentions, they probably react in a more uncontrolled way, increasing the likelihood of uncivil commenting behavior in the form of disrespect.

#### *7.3.4 Conclusions from Study IV*

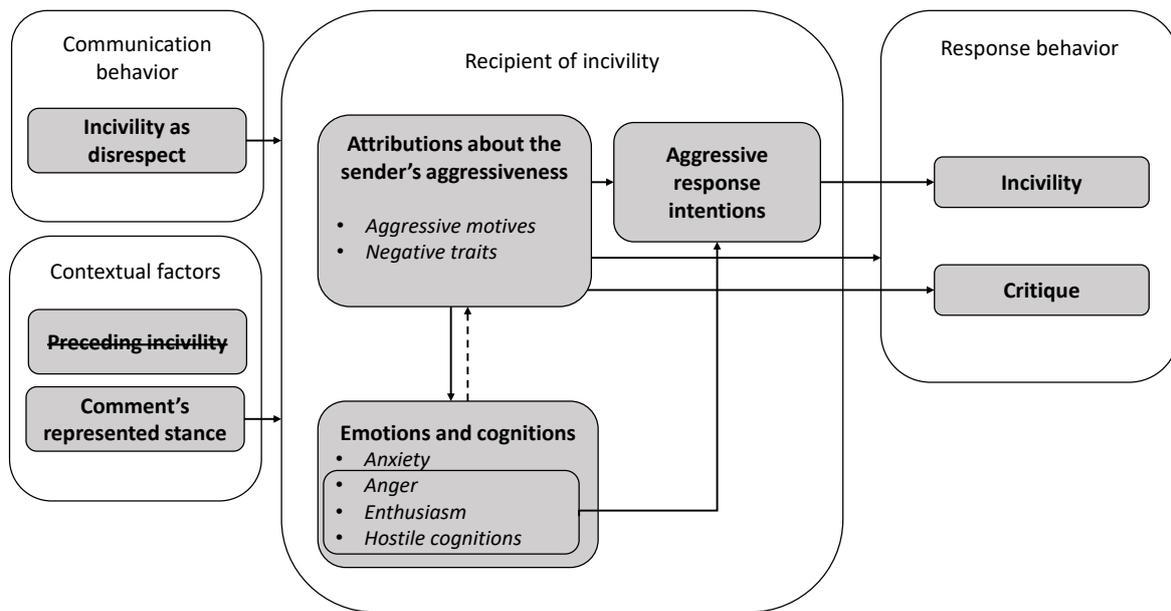
Study IV added some important insights into how incivility is processed. First, preceding incivility in other comments seems to be less critical for how people assess the discussion motives of an uncivil commenter. This does not necessarily contradict the findings of Shmargad and colleagues (2021). The authors found that people are more likely to engage in uncivil commenting behavior repeatedly when other users upvoted (i.e., affirming incivility as an injunctive norm) their initial uncivil comment or other commenters answered in an uncivil manner (i.e., affirming incivility as a descriptive norm). As these people had already behaved in an uncivil way before the other user reactions emerged, they are probably more susceptible to such norms.

Moreover, in Study IV, it was not measured whether people perceived incivility as the norm in the discussion. Therefore, it might be that people indeed considered uncivil behavior as normative in the condition with the preceding uncivil comments but still associated such behavior with an aggressive discussion motivation. This would strengthen the argumentation of Bormann and colleagues (2021), as this indicates that certain expectations for communication cannot be easily changed.

The stance of the comment was again a more important predictor for the extent of aggressive motive attributions. However, as revealed in Study III, this effect emerges independently from a comment's incivility. Nevertheless, it is an essential finding that a comment's stance can affect the attribution process, whereas the extent of incivility in other comments is less critical.

Finally, Study IV demonstrated that the interplay of emotions, cognitions, and attributions could be crucial for individuals answering intentions. Moreover, in Study II and 3, it was found that attributions of more stable characteristics (i.e., intolerable political views, a lack of discussion skills, or negative traits) increase the likelihood of critique expressions. Study IV showed that the attribution of aggressive discussion motives also triggers critique. In contrast, a negative appraisal of an uncivil commenter's discussion motives does not

automatically lead to uncivil reactions. People are only more likely to engage in such behavior when this appraisal leads to aggressive answering intentions. This corresponds to the conclusion of Study I: “It’s the aggression, stupid!” Then, as people adopt the intentions they attribute to the other side, people seemingly communicate cooperatively in an aggressive way (Allwood et al., 2000). Please see *Figure 4* for a schematic visualization of the findings from Study III and Study IV.



*Figure 4.* Illustration of the findings from Study III and Study IV. Preceding incivility is crossed out as no effects were found for this factor. Attributions influence whether people respond at all. If people respond, aggressive answer intentions determine whether people respond in an uncivil way. The arrow with the dotted line indicates the hypothetical reciprocal effect between attributions and emotions as well as cognitions.

## IV Discussion

### 8 General conclusions from the empirical studies

This dissertation aimed to investigate incivility in online political communication from a new perspective. It was claimed that incivility is not simply a message characteristic but should be considered a social act. Therefore, the perspective of social perception was assumed to be vital for explaining how people process others' uncivil commenting behavior. Here, the key research question has been: What is more critical for causes and consequences of incivility: a lack of (perceived) cooperation, (perceived) aggression, or both? It was assumed that both factors play an essential role in uncivil online (political) communication. Although results were mixed, this assumption was supported. Strengthening the theoretical assumptions of Bormann and colleagues (2021), the perspective of cooperative communication is valuable to compare and categorize different types of incivility. Further uncivil behavior seems to rise out of aggression – at least for incivility in the form of disrespect. Then, aggression seems to become the new fundament of cooperative communication. However, ideal cooperation, as described by Allwood and colleagues (2000), probably fails because mutual trust in each other's ethical actions is likely to be diminished.

Overall, it was proved that attributions about its senders are an essential factor for the processing of incivility. The following part discusses in detail what accounts for incivility as a lack of cooperation and what accounts for incivility as a manifestation of aggression, and how these two perspectives can be brought together.

#### 8.1 Incivility as a lack of cooperative communication

Based on the novel conceptualization of incivility provided by Bormann and colleagues (2021), it was argued that the common theoretical ground of different types of incivility could be found in violations of cooperative communication norms. The findings of Study I and Study II empirically supported this idea. Study I found that perceived violations of communicative norms are often the cause for further uncivil commenting. Seemingly, then, people try to guide the discussion back into normative directions (Tomasello, 2019, Chapter 9). Although these results have been descriptive due to the qualitative analysis, a relatively clear picture has emerged. Overall, participants most frequently indicated counteractive motives for behaving uncivilly in a specific case. As outlined in Chapter 4, it can have different reasons why cooperative communication fails. For instance, following Allwood and colleagues (2000), a crucial requirement for cooperation to emerge is that individuals trust each others' cooperative intentions. Likewise, they must be convinced that others cognitively

consider them. For ideal cooperation, people additionally must trust in others' ethical behavior. Although these requirements of Allwood and colleagues (2000) do not seem to be very extensive, in complex activities as in discussions, it is hard to determine on which level cooperation fails. Even when reacting in an uncivil way to perceived antinormative communication behavior, people behave cooperatively since they still cognitively consider and communicate with the "other side." Thus, a primary joint purpose (i.e., communicating with each other) is present.

Nevertheless, certain expectations about communication must have been violated. Otherwise, people would not aim to counteract the behavior, respectively. However, Study I did not allow more profound insights into why people consider others' behavior antinormative. A major shortcoming is that the study did not systematically examine which behaviors people responded to in a particular uncivil manner. Even if there is a presumption that it was other uncivil behavior, it is difficult to determine whether it was actually norm violating behavior because of the subjectivity of such appraisals. Therefore, in Study II, it was investigated how violations of communicational norms, as proposed by Bormann and colleagues (2021), are appraised by recipients in a more controlled way.

Study I revealed that individuals' assessment of others' behavior could be crucial for their uncivil behavior. In this context, it was argued that people do not simply appraise the message but rather assess the sender behind the message. Therefore, an attribution theory approach was considered appropriate to investigate how people appraise uncivil behavior in Study II. By combining the perspective of cooperative communication (Allwood et al., 2000; Bormann et al., 2021; Grice, 1975) with assumptions of attribution theory (i.e., folk-conceptual theory of behavior explanation; Malle, 1999), it was indeed found that distinct types of incivility led to different appraisals about senders' of incivility. More specifically, different types of incivility led to distinct attributions about the sender's intention to discuss, tolerable political views, and discussion skills. These findings support the thesis's assumption that a lack of cooperation can be explained on different attributional dimensions. Here, the folk-conceptual theory of behavior explanation (Malle, 1999) proved to be particularly useful, as this approach considers that people do not necessarily have a unique explanation for the behavior of others. Although Malle (1999) argues that reason explanation (i.e., the attribution of behavioral motives) has the most significant explanatory value, his framework acknowledges that other inferences can be crucial, too. When considering the prerequisites for cooperative action (Allwood et al., 2000) and the complex inferences people have to draw about others (Tomasello et al., 2005), attributions about others' behavioral motives indeed

seem not to be enough to explain (un)cooperative behavior. As incivility can take many forms, it is all the more important to understand how people process its distinct types and manifestations. In other words, it is important to know why people conclude that uncivil behavior is uncooperative. Only then, consequences of incivility can be assessed in comparison with each other.

Although the results of Study II provide valuable insights into how people assess incivility from the social perspective, the approach has not been exhaustive. Future studies should derive other cooperation-related attributions that people are likely to make when communicating with others). Moreover, this approach should be employed to investigate other types of incivility. However, it has been proven that incivility, conceptualized as a lack of cooperative communication, can help to understand different types of incivility on a common theoretical ground.

## **8.2 Incivility as aggressive communication**

Although the perspective of cooperative communication allows a valuable view on incivility in online political online communication, it does not rule out that aggression plays an essential role in uncivil commenting behavior. Indeed, Study I found that general aggressive motives for discussion are a primary driver for frequent uncivil behavior. This study also found that this effect is especially true for insulting and vulgar behavior – thus, behavior typical for disrespecting behavior. Therefore, although counteractive discussion motives are a primary driver for uncivil behavior on a single case level, aggression plays a more critical role in uncivil commenting behavior from a person-centered perspective. It also has to be noted that having counteractive motives does not exclude that people feel aggression. Indeed, earlier studies suggest that individuals have increased aggressive feelings when exposed to incivility (e.g., Chen & Lu, 2017; Gervais, 2015, 2017). That does not mean that this contradicts the assumptions about incivility and violations of cooperative communicational norms. As perceptions of uncooperative behavior and aggression are closely related (Tomasello, 2019, Chapter 9), this is instead a more nuanced perspective on the previous results. Hence, even though (a lack of) cooperative communication is the common ground that can gather different types of incivility, for disrespecting behavior, aggression might play the more crucial role – both for the processing and behaving in this way. Here, the present dissertation draws on the general aggression model (Anderson & Bushman, 2002), suggesting that individuals' appraisal of a social situation is crucial for their reaction.

Study III and Study IV drew a comparatively comprehensive picture, though not an exhaustive one, of the role of aggression in uncivil online communication in the form of

disrespecting behavior. First, by drawing on the multiple inference model (Reeder, 2009), it was demonstrated that the attribution of aggressive discussion motives is crucial for the hostile assessment of the sender of an uncivil message. Although incivility affected both the attribution of aggressive and nonaggressive discussion motives (i.e., interest-related motives and counteractive motives), only aggressive motives increased the attribution of negative traits. This finding is notable because it demonstrates that aggression rather than a perceived lack of cooperation leads to hostile thoughts about an uncivil commenter. As it was shown in Study III that such a hostile appraisal of a person diminishes recipients' willingness to join a discussion, it can be argued that, through this path, incivility has undesired consequences for the democratic exchange by impeding cooperative communication. Although this effect was small, it demonstrates that negative appraisals of others can shape discussion dynamics. The unwillingness to discuss can be considered inherently uncooperative. Even basic requirements such as the joint purpose to communicate or cognitively considering each other are not present (Allwood et al., 2000). Hence, hostile appraisals can have a detrimental effect on subsequent cooperative communication as they already prevent the initial stages of such communication. However, interestingly, such an appraisal did not increase uncivil answers. Only critique expressions were more likely when people considered an uncivil commenter to be an "idiot." This finding did not comply with the idea that incivility rises out of preceding incivility through the appraisal process. That was assumed based on the results of Study I and previous findings (Chen & Lu, 2017; Gervais, 2015; Kwon & Gruzd, 2017; Ziegele et al., 2018). However, already Study II did not reveal such a pattern. As Study I implied that people are only engaged in counteractive incivility in single cases, it was reasonable that something more is needed for people to respond in an uncivil way.

Indeed, the dissertation's last study revealed some additional insights into how attributions about an uncivil sender influence the emergence of uncivil reactions. It was shown that the likelihood of uncivil reactions only increased when aggressive attributions lead to an aggressive answering intention. Hence, only when the attribution process triggers such intentions, do people seemingly disrespectfully react to disrespect. This result closes the loop to Study I, showing that aggressive discussion motives also drive counteractive incivility. Albeit this effect was relatively small, it significantly adds to the understanding of when incivility leads to further uncivil communication.

In this connection, it was found that emotions play a crucial role in guiding behavioral intentions. Foremost, anger and enthusiasm directed which of recipients' answering intentions were more prevalent. This is in line with previous works that highlighted the function of

emotions in processing incivility (Chen & Lu, 2017; Gervais, 2015, 2017; Hutchens et al., 2019). However, Study IV added to these findings by demonstrating that anger plays an ambivalent role as it can increase both aggressive and discussion-centered (thus, cooperative) answering intentions. That finding corresponds to earlier results that incivility can also have beneficial consequences, such as increased political participation (Borah, 2014; Chen & Lu, 2017). As anger is an aversive state that people typically want to reduce (Marcus et al., 2006), recipients of uncivil messages probably engage in different strategies to neutralize this emotion. Although Study IV did not examine such effects, following the general aggression model (Anderson & Bushman, 2002), it probably depends on individuals' consciousness and self-control, whether they have aggressive answer intentions or discussion-centered answer intentions. When reacting in an uncivil manner, individuals seemingly try to reduce the aversive state by engaging in venting behavior. When having discussion-centered answering intentions, people probably aim to avoid this aversive state more constructively. However, as this has not been explicitly examined, it is on a hypothetical level. Here, more research is needed.

In sum, for disrespecting behavior, the focus on aggressive communication behavior adds some valuable insights into how incivility is processed and how it leads to further uncivil commenting behavior. From this point of view, incivility would be associated with aggression rather than a lack of cooperative communication. Nevertheless, as already noted, combining the two perspectives seems to be a promising approach. Therefore, next, it is discussed how these two approaches fit together.

### **8.3 Combined perspectives**

The present dissertation has shown that both perspectives offer added value to the understanding of how incivility is processed by individuals: Incivility as a lack of cooperative communication as well as incivility as aggressive communication. Thus, combining both views seems to be most valuable to explain how people process incivility and which consequences incivility has. As already noted, aggression is a typical strategy to punish violators of cooperative norms (Tomasello, 2019, Chapter 9). However, as people inherently expect others' communication behavior to be cooperative (Bormann et al., 2021; Grice, 1975), people must first conclude that specific behavior (e.g., uncivil commenting behavior) is uncooperative. Only when they come to such a conclusion, problems of subsequent cooperation arise from the recipient's side. Then, individuals will probably make further attributions that lead to an appraisal of the behavior because "cooperation only pays off if others cooperate as well" (Gollwitzer & Rothmund, 2009, p. 137). In other words, people

must first scan whether others' behavior in potential cooperative behavior is cooperative. Here, individuals have to draw various inferences from others' behavior (Tomasello et al., 2005). In this case, the folk-conceptual theory of behavior explanation (Malle, 1999) seems to be more helpful as it focuses on how people explain others' behavior. Thus, people make a bigger picture of the communicative situation. When they recognize failures within the other side's cooperative behavior (e.g., a lack of cooperative intentions, cooperation skills, or ethical considerations), they have to make further attributions to decide how to react to the norm violating behavior (see also Anderson & Bushman, 2002).

In the context of virtual aggression, it was already demonstrated that aggressive behavior decreases trust in others' cooperative behavior and, consequently, negatively affected cooperation with them (Rothmund et al., 2011). This finding corresponds with the argument that trust in others' cooperative behavior plays a crucial role in successful cooperation (Allwood et al., 2000). However, Tomasello (2019, Chapter 9) argues that aggression is a typical reaction to punish violators of cooperative norms. He also notes that it depends on individuals' reasons for being uncooperative (see also Kachel et al., 2018). Hence, when people have concluded that others' behavior is uncooperative, they probably make further inferences about the reasons to respond to the uncooperative behavior appropriately.

In this stage, Reeder's (2009) multiple inference model seems to be more helpful than the folk-conceptual theory of behavior explanation (Malle, 1999) to explain how people react to incivility. People then need to come to an appraisal of the behavior. As Anderson and Bushman (2002) argue, this often goes hand in hand with the evaluation of a person. Of course, Malle's (1999) approach also includes implicit appraisals about others. However, his theory focuses on the explanation process, whereas the multiple inference model (Reeder, 2009) explicitly addresses the process of how people come to more stable inferences about others. In the case of incivility, the present dissertation showed that this approach is indeed helpful to understand how people process incivility as aggressive-motivated behavior. Although not explicitly investigated in the current work, based on the above arguments, it can be assumed that this process is secondary to the conclusions about cooperative behavior.

In this dissertation, this subordinate process was only examined in the context of incivility in the form of disrespect. This type of incivility seemed particularly important as it has crucially inhibited the perception of cooperation at various attributional levels in Study II. In future studies, it should be examined which motives people attribute to senders of other types of incivility. For instance, it is unlikely that individuals attribute aggressive discussion intentions to individuals who deviate from the discussion topic. Nevertheless, such behavior

can increase aversive emotions and, therefore, might also lead to aggressive response reactions.

The empirical investigations of the current thesis also revealed that certain factors influence the usage of incivility and its processing. In accordance with the general aggression model (Anderson & Bushman, 2002), it was focused on individuals' personality and contextual factors. Surprisingly, in Study IV, we found that preceding incivility does not affect peoples' appraisal of an uncivil comment. However, individuals' dark side of personality seems to influence their uncivil commenting behavior and also can affect the processing. Moreover, the stance towards the discussion topic can bias attributions when processing uncivil messages, though no interaction effect emerged. Although these results primarily refer to the view of aggressive communication, from a theoretical point of view, it is also to be expected that these factors influence the evaluation of cooperative communication.

### ***8.3.1 The role of personality***

Study I demonstrated that personality plays an essential role in the usage of incivility. Foremost, "dark" personality traits like psychopathy and Machiavellianism seem to increase aggressive discussion motives and, in this way, foster uncivil commenting behavior. Likewise, in Study III, it was found that psychopathy can lever out the impact of incivility on the attribution of aggressive discussion motives. However, compared to the effects seen in Study I, these were comparatively small. Nonetheless, these results suggest that individuals' personality traits are important for uncivil comment behavior and its processing. One shortcoming of the current dissertation is that it was not investigated how these traits affect cooperative behavior and perceptions of cooperation. Here, psychopathy and Machiavellianism have already been found to impede cooperative behavior (Malesza, 2020). The lack of social cognitive abilities that can be found in individuals high in these traits (Ali et al., 2009; Ali & Chamorro-Premuzic, 2010) is probably the main reason for that. As cooperation requires a broad spectrum of social cognitions (Tomasello et al., 2005), people high in psychopathy and Machiavellianism probably fail to „read others' minds" adequately.

Yet, the dark triad is only one of many personality systems that can be related to uncivil online commenting. As already noted, the extent to which people are conflict-orientated might be crucial for using and processing incivility (Sydnor, 2019). Moreover, the impact of Big Five personality traits in this context has been investigated by only a few studies so far (e.g., Kenski et al., 2020). Concerning cooperative behavior, individuals' sensitivity to mean intentions can affect trust in others' cooperative behavior (Gollwitzer & Rothmund, 2009). Thus, much more work is needed to understand the role of personality

within the current approach. Nonetheless, it can be stated that an individual's personality can influence whether individuals behave in an uncivil way and how they process incivility. In the current work, this was shown empirically for the dark side of peoples' personality in the context of incivility as aggressive communication. From a theoretical perspective, this should also apply to the perspective of incivility as uncooperative communication behavior.

### ***8.3.2 The role of comment's stance***

Based on assumptions about patterns of social categorization (Turner et al., 1987), Study III and Study IV demonstrated that the represented stance towards the discussion topic, represented in comments, biases how recipients process it. Adding to previous findings (Chen & Lu, 2017; Gervais, 2015, 2017; Hutchens et al., 2019), these are the first insight into how disagreement can affect attributions about uncivil commenters. Foremost, attributions of aggressive discussion motives were increased by disagreement. Surprisingly, Study III found that a commenter's stance did not influence how incivility is processed. However, disagreement caused similar, albeit weaker, effects as incivility. Hence, it can be viewed as an amplifier for the adverse impact of incivility on attribution processes. That pattern was also confirmed by Study IV.

These findings provide some important implications. First, independently from the discussion about the consequences of incivility, they indicate that we have seemingly not yet reached the point where people regard others with different opinions as entirely legitimate adversaries, as stated by Mouffe (2000). People probably see others still as antagonists since they associate increased aggressive discussion motives with an opposing stance. From this perspective, people still have to be educated that disagreement is not necessarily related to aggressive intents. Again, the effects of a comment's represented stance have only been investigated in the context of incivility as aggressive communication. However, as with the influence of personality traits, similar effects can likely be found regarding (un)cooperative communication. Social categorization (i.e., perceptions of out-group in-group differences) can also lead to biases in cooperative actions because people are likely to have decreased trust in others' cooperative behavior (Kramer et al., 2001). Thus, disagreement can probably reduce trust in others' cooperative discussion behavior. Here, future studies should investigate what patterns emerge in combination with incivility.

## **9 Theoretical implications**

The empirical findings of the thesis validate the usefulness of the social perspective on incivility. Based on the findings and theoretical assumptions discussed in Chapter 8, it can be modeled how individuals process incivility as a social act. It was also shown that this

approach discloses a new way to categorize incivility based on recipients' appraisals about its sender. Therefore, it empirically validates and adds to the theoretical typology of incivility introduced by Bormann and colleagues (2021). Finally, the thesis provides some theoretical implications for attribution theory and its application.

### **9.1 Modeling the social processing of incivility**

Based on the studies' empirical findings and the related theoretical discussion, a model can be derived that describes how people process incivility as a social act (please see *Figure 5* below). For a better overview, the results and theoretical assumptions discussed in Chapter 8 are summarized in the following to describe the model's integral parts:

Study I has shown that the main drivers for uncivil commenting behavior are aggressive or counteractive discussion motives. It was demonstrated that individuals' personality traits substantially impact the occurrence of aggressive discussion motives. Although Study I focused on the individuals' dark side of personality, the influence of other traits (e.g., the Big Five personality traits) is very likely.

If people act in an uncivil manner, not because of aggressive motives but because they have the intention to counteract others' communicative behavior, they must have concluded that this behavior is not normative. The findings of Study II to IV provide considerable insight into how people come to such a conclusion and allow us to model this process more generically. Here, the novel approach to uncivil communication by Bormann and colleagues (2021), the general aggression model (Anderson & Bushman, 2002), and attribution theories (Malle, 1999; Reeder, 2009) serve as a vital theoretical fundament.

Study II found that uncivil communication affects attributions about the sender's cooperativeness. In turn, these attributions partly predicted critique expressions. However, it was not confirmed that attributions about uncivil commenters trigger subsequent uncivil commenting behavior – but results implied that certain types of incivility lead to further incivility per se. This was found for incivility in the form of deviating behavior, but this is probably also true for other types of incivility not examined in the study (Study II).

As discussed above, it is likely that attributions about others' cooperativeness also interact with other inferences from a theoretical perspective. Study III and IV demonstrated that attributions about aggressiveness and adverse traits are particularly important when appraising disrespectful behavior. Yet, the processes examined in these studies are probably not exclusively applicable for this type of incivility but can also be transferred to other types. For instance, violating democratic norms might trigger similar patterns when manifested in threatening behavior.

Study IV also revealed that the interplay of emotions, cognitions, and attributions plays an essential role in processing uncivil messages. Although this was only investigated for the attribution of aggressive discussion motives in the current thesis, following the general aggression model (Anderson & Bushman, 2002), this should also apply to other (more favorable) attributions.

Study III focused again on individuals' dark side concerning the influence of recipients' personality traits on the appraisal process. It was shown that psychopathy affects the attributions of aggressive discussion motives. However, as an individual's personality is very complex, many other traits potentially affect this process (e.g., Gollwitzer & Rothmund, 2009; Kenski et al., 2020; Sydnor, 2019). Therefore, a more generic influence of people's personality traits can be assumed.

Moreover, Study III and Study IV demonstrated that the stance of (uncivil) messages is a significant, influential variable. Thus, people attribute less desirable motives and traits to people with an opposing stance. As discussed in Chapter 7.1, this should be true for most kinds of attributions.

Finally, the process including emotions, cognitions, and distinct attributions about an uncivil sender's aggressiveness and cooperativeness determines how individuals respond. This can be refraining from commenting (Study III), but also the expression of critique (Study II, III, and IV) or uncivil reactions (Study IV). However, whether people respond in an uncivil manner depends on the extent of aggressive answering intentions (Study IV). Therefore, as illustrated in Figure 5, if recipients' aggressive response intentions increase, they can become the sender of an uncivil message. Then, seemingly aggressive communication is the new cooperative communication.

In sum, the findings of the current dissertation offer vital evidence for the derived model. Nevertheless, further research is needed to confirm these results and empirically validate the theoretical assumptions included in the model. Of course, the model is not exhaustive. Yet, it provides an essential first step to explaining how people process and respond to incivility from a social perspective. Although the model was developed and examined in the online context, it is not necessarily limited to this context as it draws on fundamental social psychological assumptions. Therefore, future research should examine the validity of the model in other communication contexts.

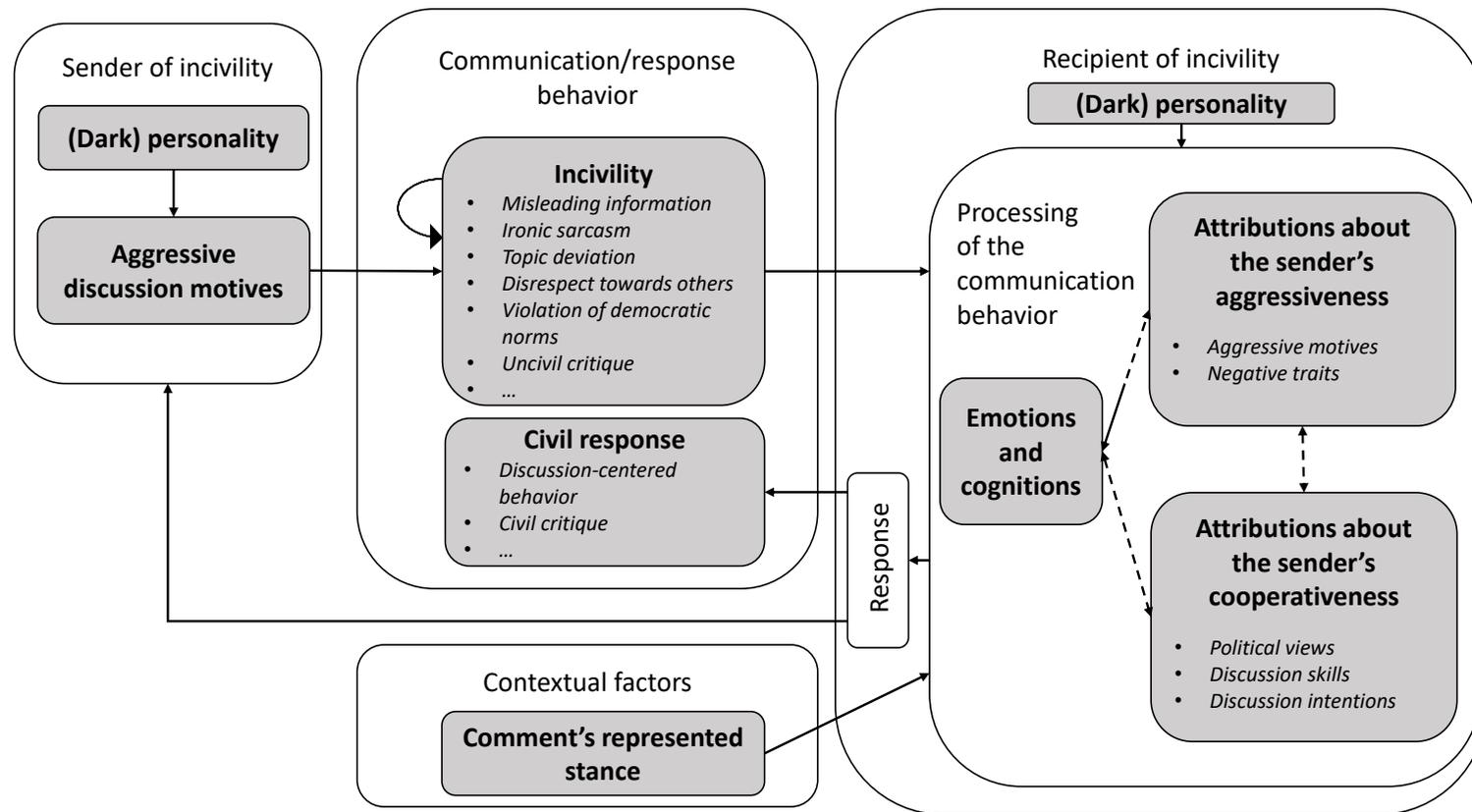


Figure 5. Model describing the social processing of incivility. Dotted lines indicate theoretical assumptions that have not been empirically investigated. The arrow pointing from incivility back to incivility indicates that certain types can facilitate further uncivil communication by themselves (e.g., deviating behavior).

## 9.2 New opportunities to categorize incivility

Further theoretical implications arise from the opportunity to categorize incivility based on recipients' appraisal: First, the results of Study II can complement the theoretically derived typology introduced by Bormann and colleagues (2021). The authors argue that incivility can cause different problems regarding cooperative actions: problems of motivation (e.g., prioritizing individual goals over joint goals), cognition (i.e., insufficient common knowledge), and coordination (i.e., unclear or ambiguous responsibilities). However, the authors do not assume that different problems are more or less likely to occur depending on the type of incivility. Instead, they argue that incivility can lead to problems on all levels. Study II complemented this approach by focusing on the recipient's perspective. In this way, it could be empirically confirmed that problems as described by Bormann and colleagues (2021) can arise. The study's findings imply that specific problems occur more likely than others based on the recipients' attributions. For example, it was shown that topic deviating behavior or disrespect towards others lead more likely to problems on a motivational level because people attribute particular low levels of discussion intentions to the sender.

Second, the approach allows categorizing different types of incivility in a recipient-centered way. As incivility lies in the eye of the beholder (Chen et al., 2019; Herbst, 2010), it offers a new way of understanding how different types of incivility affect its recipients. So far, on a perceptual level, different types of incivility have only been categorized by the extent to which people consider an expression to be uncivil (Kenski et al., 2020; Stryker et al., 2016). Such approaches have different shortcomings. Foremost, it is a unidimensional perspective on a multidimensional concept. A message can, of course, be considered to be uncivil, but at the same time, people can regard such a behavior to be appropriate (e.g., when verbally punishing another person). Hence, this assessment says little about the impact of uncivil messages on the recipients. Likewise, assessing a specific communicative behavior to be uncivil says little about its evaluation in the context of concrete communication. For instance, Stryker and colleagues (2016) found that deceptive behavior is regarded as uncivil by people but Study II did not find that spreading misleading information led to a negative appraisal of the sender. It is therefore reasonable that this behavior is considered uncivil in the abstract but has little influence on processing the corresponding message in concrete situations.

The social perspective introduced by the current dissertation allows a multidimensional categorization of distinct types of incivility based on recipients' appraisals. Results of Study II confirmed that people make different attributions about commenters

depending on the type of incivility they use. Therefore, it is possible to examine more precisely how the different types of incivility influence the dynamics of discussions. Interestingly, confirming previous studies (Kenski et al., 2020; Stryker et al., 2016), Study II found that incivility in the form of disrespecting behavior is the most problematic. However, in addition to former approaches, it was found that disrespecting behavior severely impedes cooperation on different dimensions. Such behavior seems to affect cooperative behavior much more substantially than other types of incivility.

Taken together, this dissertation complements the framework of Bormann and colleagues' (2021). It provides the first empirical evidence for their typology and additionally provides recipient-centered subcategorization of the distinct types of incivility. The usefulness of other attributions (e.g., aggressive discussion motives) to categorize incivility should be investigated within the next steps.

### **9.3 The utility of attribution theories**

The present doctoral thesis also provides some theoretical implications for the application of attribution theories. Although attribution-making plays a crucial role in other theoretical frameworks like the hypothesis of shared intentionality introduced by Tomasello and colleagues (2005) or the general aggression model (Anderson & Bushman, 2002), comparatively little attention is paid to this aspect within these approaches. However, the folk-conceptual theory of behavior explanation (Malle, 1999) and the multiple inference model (Reeder, 2009) show how complex this process is. Hence, attributions theories should be stronger integrated when modeling how people interpret others' behavior. Especially the general aggression model (Anderson & Bushman, 2002) seems to be a beneficial supplement since the model explicitly refers to attribution theory when describing how people come to an appraisal about an event. Therefore, integrating concrete assumptions about the attribution process would make the model more comprehensible. As Reeder's (2009) multiple inference model describes how people come to appraisals about others, this attribution theory seems to be central in particular here.

This also leads to the next theoretical implication. As discussed in Chapter 3.3, some attribution theories might be more helpful than others in certain situations. Basically, it can be distinguished between attribution theories that focus on how individuals explain others' behavior and theories that focus on the inferences of dispositions (e.g., traits) people make from others' behavior (Malle & Korman, 2013). So far, little effort has been made to bring these two approaches together. The results of the current dissertation suggest that different attributions might be made within different appraisal stages. For instance, when people

appraise others' cooperative behavior, they have to explain the interaction partner's behavior to ensure that cooperation is successful (see also Tomasello et al., 2005). As long as cooperation succeeds, people do not necessarily make attributions about others' stable characteristics. However, when cooperation fails, people must come in some ways to an appraisal of the person to decide how to react to the norm violating behavior. Then dispositional attribution theories like the multiple inference model (Reeder, 2009) seem to be more helpful to explain how people reach such an assessment. Of course, these two approaches cannot be sharply distinguished, and the assumptions are on a rather speculative basis. Nevertheless, future studies should address in which situation which attribution theories are more applicable and how the different approaches are related.

## **10 Practical implications**

The four studies conducted within the dissertation also provide some practical implications for research and administrators of online discussion boards. Foremost, people's attributions when exposed to incivility offer some ideas on dealing with uncivil discussion behavior in the future.

### **10.1 Measuring incivility**

The approach presented in this dissertation not only enables new ways of categorizing incivility (see Chapter 4.2). It also provides new ways of how incivility can be measured on a perceptual level. People's attributions about senders of uncivil messages allow for a multidimensional assessment of incivility beyond its mere manifestation. As shown in Study II to IV, such a measurement assesses whether people regard a behavior as uncooperative or aggressive and considers why people do that. Primarily Study II provides a more comprehensive way to measure individuals' perception of uncivil online behavior through the lens of cooperation. Results showed that this approach is vital for the understanding of differences between distinct types of incivility. Therefore, further scale-based measurements of perceived incivility should consider the different attributional dimensions assessed in Study II or expand this approach. Another benefit is that the assessed attributions can be customized in each case depending on the research focus (e.g., aggressive vs. cooperative commenting motives). However, a more substantial scale development is needed. Although the scales used in the studies have consistently shown good reliability, more comprehensive methods for validation are required. The present work shows that it is worthwhile to develop standardized measurements that include different attributions for further research on online incivility.

## 10.2 Interventions

The findings of the dissertation also give some implications for designing interventions that counter incivility. A vital benefit of the dissertation's approach is that it makes it easier to identify harmful types of incivility. For instance, in Study II, it was shown that incivility in the form of disrespect is crucially assessed as uncooperative on distinct dimensions. Therefore, such a type of incivility probably needs to be countered with more vigor and through more measures than other types. Moreover, by assessing distinct dimensions of attributions, interventions against certain types of incivility can be better tailored. As this approach arises from the perspective of social perception, implications for social interventions are the most obvious.

Concerning the recipients of incivility, people should be encouraged to reconsider the attributions they have made to prevent an immediate disruption of the discussion. For instance, when exposed to topic deviations, people could be informed that the sender can still have the intention to discuss and that the person should be guided back to the topic. Likewise, it might help to tell people that the sender does not necessarily lack discussion skills when exposed to irony or sarcasm. They could be encouraged to tell the norm violator that they associate a lack of discussion skills with such behavior. Of course, this does not prevent incivility from appearing in discussions. However, such interventions might help avoid misunderstandings in communication and strengthen discussion dynamics by showing users ways to cope with certain types of incivility. This approach seems to be promising, as only people with aggressive response intentions were inclined to respond disrespectfully. Thus, by triggering a reappraisal as proposed by the general aggression model (Anderson & Bushman, 2002), people might be encouraged to answer in a discussion-centered manner. This can also be important when people are confronted with disagreement. As an opposing stance can negatively bias the attribution process independently of incivility, people should also be aware that they are susceptible to such effects to avoid additional undesired consequences for discussion dynamics.

Of course, the question arises how such measures can be implemented. Here, different approaches are conceivable. For example, this could be a supplement to auto-detection intervention. Before deleting undesired comments, other discussants could be informed as described above by pop-up messages. Likewise, discussion board moderators or guidelines (e.g., netiquettes) could share such information, whereas moderators could do this more targeted. However, most important is probably that such information is included in media

literacy campaigns (Arendt, 2013) that focus on online discussions to educate people in the long term.

Regarding senders of incivility, similar approaches might be beneficial. For instance, senders of disrespectful messages or sarcastic individuals could be informed that others regard their behavior as unskillful or generally negatively appraise them. Then, feelings of cognitive dissonance (Festinger, 1957) may arise, which leads a sender to refrain from this behavior. Likewise, individuals could be informed that uncivil communication impedes cooperative communication even when it is not intended. However, discussion board administrators should be aware that individuals who frequently discuss in an uncivil manner are probably driven by a dark personality and, therefore, are assumably harder to influence by simple educational approaches. Consequently, in these cases, the exclusion of the persons from the discussion might be justifiable.

### **11 Limitations and future studies**

Although it has been shown that this is a promising approach, some overarching limitations must be mentioned. First, the generalizability of the findings is limited in certain ways: In all studies, the sample was not representative as we draw on convenience samples. As a result, participants have been rather female, well educated, and left-leaning. This imbalance might have biased the results. Foremost the perception of incivility seems to be influenced by these variables. For instance, Kenski and colleagues (2020) found that females are more likely to assess potentially uncivil speech as actually uncivil. In contrast, conservatives were less likely to do so. However, as participants were recruited most on social media, it was accomplished to reach people who discuss online. Especially in Study I and Study II, a comparatively large number of people participated who comment online regularly (cf. Springer et al., 2015).

Further, all studies addressed a German-speaking population. It still has to be investigated whether the findings of the current dissertation apply to other cultural backgrounds. It was already indicated that people's cultural background affects how people use social media platforms (Gil de Zúñiga et al., 2017), how they react to uncivil behavior (Brauer & Chaurand, 2009), and which attributions they make (Betancourt et al., 1992). Therefore, future studies should expand the approach of the present work to other cultural contexts.

Another limitation of the generalizability of the findings results from the operationalization of incivility in the four studies. Although care was taken to use different manifestations of incivility in each study, uncivil commenting behavior can theoretically

show up in an infinite number of ways (see also Chen, 2017). Varying operationalizations of the different types of incivility might cause different results. Future research must validate the findings for other manifestations of incivility.

Some further limitations result from the methodological approaches employed in the four studies. In Study III and IV, participants have been exposed to hypothetical scenarios. Therefore, the external validity of the findings is limited. Likewise, although in Study II, participants were told that they comment in an allegedly real discussion forum, it had uncommon discussion rules impairing the external validity. Hence, future studies should investigate the influence of incivility on attribution processes in more realistic settings.

It also has to be noted that all measures relied on individuals' self-reports. Likewise, in Study II to IV, uncivil commenters had been assessed on preselected items. Thus, it is not given that people make exactly these attributions. Future studies should examine other potential attributions people make when being exposed to incivility. Here more open and implicit measures could be beneficial. For instance, diary-based methods might help explore which attributions people make when reading online discussions.

Finally, as the dissertation focused on the perceptual level of incivility, it was hardly examined which undesirable consequences distinct types of incivility can have beyond the appraisal of the recipients. Foremost, the violation of democratic norms has much potential to do lasting harm to our democracy (Papacharissi, 2004). Therefore, the present approach must be delimited from the research of such consequences.

## 12 Conclusion

This dissertation aimed to strengthen the perspective of incivility as a social action instead of a mere message characteristic. Based on four empirical studies and in-depth theoretical reasoning, a model was derived that explains how incivility arises and is processed from the perspective of social perception. At its heart, the model builds on assumptions about cooperation (Allwood et al., 2000; Bormann et al., 2021; Grice, 1975; Tomasello et al., 2005), attribution theories (Malle, 1999; Reeder, 2009), and the general aggression model (Anderson & Bushman, 2002). In this way, an integrative approach was introduced that allows investigating incivility in online political discussions from a social perspective. Core findings are that incivility is not only triggered by aggressive discussion motives but can also be part of counteractions against preceding norm violations. It was also found that incivility leads to more negative attributions about its sender. This includes attributions about the sender's cooperativeness and aggressiveness. However, if such attributions lead to further incivility seemingly depends on whether they develop aggressive response intentions or not. Thus,

interventions against incivility should take the most care that people do not come to such behavioral intentions to avoid spirals of incivility.

Supporting the recently introduced approach by Bormann and colleagues (2021), it was shown that different types of incivility lead to varying attributions regarding the sender's suitability as a cooperative discussion partner. In this way, it was confirmed that incivility causes certain problems that potentially impede cooperative communication. Furthermore, the differences that have been found regarding recipients' attribution making allow for subcategorization and new ways of measuring the different types of incivility on a perceptual level.

Taken together, looking at incivility from the social perspective has proven to be an added value for this field of research. In particular, attribution theories have proven to be a useful framework to understand how people process incivility. Future research should more strongly include this view and should expand the approach of the current dissertation.

**V References**

- Ajzen, I. (1987). Attitudes, Traits, and Actions: Dispositional Prediction of Behavior in Personality and Social Psychology. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 20, pp. 1–63). Academic Press. [https://doi.org/10.1016/S0065-2601\(08\)60411-6](https://doi.org/10.1016/S0065-2601(08)60411-6)
- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179–211. [https://doi.org/10.1016/0749-5978\(91\)90020-T](https://doi.org/10.1016/0749-5978(91)90020-T)
- Ali, F., Amorim, I. S., & Chamorro-premuzic, T. (2009). Empathy deficits and trait emotional intelligence in psychopathy and Machiavellianism. *Personality and Individual Differences*, 47(7), 758–762. <https://doi.org/10.1016/j.paid.2009.06.016>
- Ali, F., & Chamorro-Premuzic, T. (2010). Investigating Theory of Mind deficits in nonclinical psychopathy and Machiavellianism. *Personality and Individual Differences*, 49(3), 169–174. <https://doi.org/10.1016/j.paid.2010.03.027>
- Allen, J. J., Anderson, C. A., & Bushman, B. J. (2018). The General Aggression Model. *Current Opinion in Psychology*, 19, 75–80. <https://doi.org/10.1016/j.copsyc.2017.03.034>
- Allwood, J., Traum, D., & Jokinen, K. (2000). Cooperation, dialogue and ethics. *International Journal of Human-Computer Studies*, 53(6), 871–914. <https://doi.org/10.1006/ijhc.2000.0425>
- Alonzo, M., & Aiken, M. (2004). Flaming in electronic communication. *Decision Support Systems*, 36(3), 205–213. [https://doi.org/10.1016/S0167-9236\(02\)00190-2](https://doi.org/10.1016/S0167-9236(02)00190-2)
- 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, A. A., Yeo, S. K., Brossard, D., Scheufele, D. A., & Xenos, M. A. (2018). Toxic Talk: How Online Incivility Can Undermine Perceptions of Media. *International Journal of Public Opinion Research*, 30(1), 156–168. <https://doi.org/10.1093/ijpor/edw022>
- Anderson, C. A., & Bushman, B. J. (2002). Human Aggression. *Annual Review of Psychology*, 53(1), 27–51. <https://doi.org/10.1146/annurev.psych.53.100901.135231>
- Arendt, F. (2013). Dose-Dependent Media Priming Effects of Stereotypic Newspaper Articles on Implicit and Explicit Stereotypes. *Journal of Communication*, 63(5), 830–851. <https://doi.org/10.1111/jcom.12056>
- Barnidge, M. (2015). The role of news in promoting political disagreement on social media. *Computers in Human Behavior*, 52, 211–218. <https://doi.org/10.1016/j.chb.2015.06.011>

- Barnidge, M. (2017). Exposure to Political Disagreement in Social Media Versus Face-to-Face and Anonymous Online Settings. *Political Communication*, 34(2), 302–321. <https://doi.org/10.1080/10584609.2016.1235639>
- Barnidge, M., & Rojas, H. (2014). Hostile media perceptions, presumed media influence, and political talk: Expanding the corrective action hypothesis. *International Journal of Public Opinion Research*, 26(2), 135–156. <https://doi.org/10.1093/ijpor/edt032>
- Benkler, Y. (2006). *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. Yale University Press.
- Betancourt, H., Hardin, C., & Manzi, J. (1992). Beliefs, Value Orientation, and Culture in Attribution Processes and Helping Behavior. *Journal of Cross-Cultural Psychology*, 23(2), 179–195. <https://doi.org/10.1177/0022022192232004>
- Blom, R., Carpenter, S., Bowe, B. J., & Lange, R. (2014). Frequent Contributors Within U.S. Newspaper Comment Forums. *American Behavioral Scientist*, 58(10), 1314–1328. <https://doi.org/10.1177/0002764214527094>
- 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*. Advance online publication. <https://doi.org/10.1093/ct/qtab018>
- Boulianne, S. (2015). Social media use and participation: a meta-analysis of current research. *Information Communication and Society*, 18(5), 524–538. <https://doi.org/10.1080/1369118X.2015.1008542>
- Boulianne, S. & Koc-Michalska, K. (2021). The Role of Personality in Political Talk and Like-Minded Discussion. *The International Journal of Press/Politics*, 27(1), 285–310. <https://doi.org/10.1177/1940161221994096>
- Boyd, R. (2004). Michael Oakeshott on Civility, Civil Society and Civil Association. *Political Studies*, 52(3), 603–622. <https://doi.org/10.1111/j.1467-9248.2004.00498.x>
- Brauer, M., & Chaurand, N. (2009). Descriptive norms, prescriptive norms, and social control: An intercultural comparison of people’s reactions to uncivil behaviors. *European Journal of Social Psychology*, 40(3), 490–499. <https://doi.org/10.1002/ejsp.640>
- Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge University Press.

- Brundidge, J. (2010). Encountering “Difference” in the contemporary public sphere: The contribution of the internet to the heterogeneity of political discussion networks. *Journal of Communication, 60*(4), 680–700. <https://doi.org/10.1111/j.1460-2466.2010.01509.x>
- Buckley, K. E., & Anderson, C. A. (2006). A theoretical model of the effects and consequences of playing video games. In P. Vorderer & J. Bryant (Eds.), *Playing Video Games: Motives, Responses, and Consequences* (pp. 363–378). Lawrence Erlbaum Associates. <https://doi.org/10.1111/j.1460-2466.2006.00323.24>
- Ceron, A., & Memoli, V. (2016). Flames and Debates : Do Social Media Affect Satisfaction with Democracy ? *Social Indicators Research, 126*(1), 225–240. <https://doi.org/10.1007/s11205-015-0893-x>
- Chen, G. M. (2017). *Online incivility and public debate: Nasty talk*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-56273-5>
- Chen, G. M., Fadnis, D., & Whipple, K. (2020). Can We Talk About Race? Exploring Online Comments about Race-Related Shootings. *Howard Journal of Communications, 31*(1), 35–49. <https://doi.org/10.1080/10646175.2019.1590256>
- 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 + Society, 5*(3). <https://doi.org/10.1177/2056305119862641>
- Chen, G. M., & Ng, Y. M. M. (2016). Third-person perception of online comments: Civil ones persuade you more than me. *Computers in Human Behavior, 55*, 736–742. <https://doi.org/10.1016/j.chb.2015.10.014>
- 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>
- Cheung, C. M. K., Chiu, P., & Lee, M. K. O. (2011). Online social networks: Why do students use facebook? *Computers in Human Behavior, 27*(4), 1337–1343. <https://doi.org/10.1016/j.chb.2010.07.028>
- 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. *Advances in experimental social psychology, 24*(20), 201–234. [https://doi.org/10.1016/S0065-2601\(08\)60330-5](https://doi.org/10.1016/S0065-2601(08)60330-5)

- Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology*, 58(6), 1015–1026. <https://doi.org/10.1037/0022-3514.58.6.1015>
- 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>
- Coleman, S., & Moss, G. (2012). Under Construction: The Field of Online Deliberation Research. *Journal of Information Technology and Politics*, 9(1), 1–15. <https://doi.org/10.1080/19331681.2011.635957>
- Cooper, J., & Mackie, D. (1983). Cognitive dissonance in an intergroup context. *Journal of Personality and Social Psychology*, 44(3), 536–544. <https://doi.org/10.1037/0022-3514.44.3.536>
- Correa, T., Hinsley, A. W., & de Zúñiga, H. G. (2010). Who interacts on the Web?: The intersection of users' personality and social media use. *Computers in Human Behavior*, 26(2), 247–253. <https://doi.org/10.1016/j.chb.2009.09.003>
- Craker, N., & March, E. (2016). The dark side of Facebook: The Dark Tetrad, negative social potency, and trolling behaviours. *Personality and Individual Differences*, 102, 79–84. <https://doi.org/10.1016/j.paid.2016.06.043>
- 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>
- Dolan, R., Conduit, J., Fahy, J., & Goodman, S. (2016). Social media engagement behaviour: a uses and gratifications perspective. *Journal of Strategic Marketing*, 24(3–4), 261–277. <https://doi.org/10.1080/0965254X.2015.1095222>
- Eginli, A. T., & Tas, N. O. (2018). Interpersonal Communication in Social Networking Sites: An Investigation in the Framework of Uses and Gratification Theory. *Online Journal of Communication and Media Technologies*, 8(2), 81–104. <https://doi.org/10.12973/ojcm/2355>
- Festinger, L. (1957). An introduction to the theory of dissonance. *A Theory of Cognitive Dissonance*. <https://doi.org/10.1037/10318-001>
- Foulkes, L., Viding, E., McCrory, E., & Neumann, C. S. (2014). Social Reward Questionnaire (SRQ): Development and validation. *Frontiers in Psychology*, 5, 1–8. <https://doi.org/10.3389/fpsyg.2014.00201>

- Frischlich, L., Schatto-Eckrodt, T., Boberg, S., & Wintterlin, F. (2021). Roots of Incivility: How Personality, Media Use, and Online Experiences Shape Uncivil Participation. *Media and Communication*, 9(1), 195–208. <https://doi.org/10.17645/mac.v9i1.3360>
- Garton Ash, T. (2016). *Free Speech: Ten Principles for a Connected World*. Yale University Press.
- 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 and 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>
- Gil de Zúñiga, H., Diehl, T., Huber, B., & Liu, J. (2017). Personality Traits and Social Media Use in 20 Countries: How Personality Relates to Frequency of Social Media Use, Social Media News Use, and Social Media Use for Social Interaction. *Cyberpsychology, Behavior, and Social Networking*, 20(9), 540–552. <https://doi.org/10.1089/cyber.2017.0295>
- Gollwitzer, M., & Rothmund, T. (2009). When the need to trust results in unethical behavior: The sensitivity to mean intentions (SeMI) model. In D. De Cremer (Ed.), *Psychological perspectives on unethical behavior and decision making* (pp. 135–152). Information Age.
- Gonçalves, J. (2018). Aggression in news comments: how context and article topic shape user-generated content. *Journal of Applied Communication Research*, 46(5), 604–620. <https://doi.org/10.1080/00909882.2018.1529419>
- Goyanes, M., Borah, P., & Gil de Zúñiga, H. (2021). Social media filtering and democracy: Effects of social media news use and uncivil political discussions on social media unfriending. *Computers in Human Behavior*, 120. <https://doi.org/10.1016/j.chb.2021.106759>
- Graham, T. (2015). Everyday political talk in the Internet-based public sphere. In S. Coleman & D. G. Freelon (Eds.), *Handbook of Digital Politics* (pp. 247–263). Edward Elgar Publishing. <https://doi.org/10.4337/9781782548768.00024>
- Grice, H. P. (1975). Logic and Conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax And*

- Semantics* (3rd ed., Issue January, pp. 41–58). Academic Press.
- Groshek, J., & Cutino, C. (2016). Meaner on Mobile: Incivility and impoliteness in communicating online. *Proceedings of the 7th 2016 International Conference on Social Media & Society - SMSociety '16, 1991*, 1–7. <https://doi.org/10.1145/2930971.2930976>
- Halpern, D., & Gibbs, J. (2013). Social media as a catalyst for online deliberation? Exploring the affordances of Facebook and YouTube for political expression. *Computers in Human Behavior*, 29(3), 1159–1168. <https://doi.org/10.1016/j.chb.2012.10.008>
- 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>
- Han, S.-H., Brazeal, L. M., & Pennington, N. (2018). Is Civility Contagious? Examining the Impact of Modeling in Online Political Discussions. *Social Media + Society*, 4(3). <https://doi.org/10.1177/2056305118793404>
- Heckhausen, J. (2000). Evolutionary Perspectives on Human Motivation. *American Behavioral Scientist*, 43(6), 1015–1029. <https://doi.org/10.1177/00027640021955739>
- Heckhausen, J., & Heckhausen, H. (2018). *Motivation and Action* (3rd ed.). Springer International Publishing. <https://doi.org/10.1007/978-3-319-65094-4>
- Heider, F. (1958). *The Psychology of Interpersonal Relations*. Lawrence Erlbaum. <https://doi.org/10.2307/2089062>
- Herbst, S. (2010). *Rude Democracy: Civility and Incivility in American Politics*. Temple University Press.
- Hogg, M. A., & Reid, S. A. (2006). Social identity, self-categorization, and the communication of group norms. *Communication Theory*, 16(1), 7–30. <https://doi.org/10.1111/j.1468-2885.2006.00003.x>
- Hutchens, M. J., Silva, D. E., Hmielowski, J. D., & Cicchirillo, V. J. (2019). What's in a username? Civility, group identification, and norms. *Journal of Information Technology & Politics*, 16(3), 203–218. <https://doi.org/10.1080/19331681.2019.1633983>
- Hwang, H., Kim, Y., & Huh, C. U. (2014). Seeing is Believing: Effects of Uncivil Online Debate on Political Polarization and Expectations of Deliberation. *Journal of Broadcasting & Electronic Media*, 58(4), 621–633. <https://doi.org/10.1080/08838151.2014.966365>
- 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>

- Hwang, H., Pan, Z., & Sun, Y. (2008). Influence of Hostile Media Perception on Willingness to Engage in Discursive Activities: An Examination of Mediating Role of Media Indignation. *Media Psychology, 11*(1), 76–97.  
<https://doi.org/10.1080/15213260701813454>
- Jamieson, K. H., & Hardy, B. (2012). What Is Civil Engaged Argument and Why Does Aspiring to It Matter? *PS: Political Science & Politics, 45*(3), 412–415.  
<https://doi.org/10.1017/S1049096512000479>
- Jones, E. E., & Davis, K. E. (1965). From Acts To Dispositions The Attribution Process In Person Perception. *Advances in Experimental Social Psychology, 2*, 219–266.  
[https://doi.org/10.1016/S0065-2601\(08\)60107-0](https://doi.org/10.1016/S0065-2601(08)60107-0)
- Kachel, U., Svetlova, M., & Tomasello, M. (2018). Three-Year-Olds' Reactions to a Partner's Failure to Perform Her Role in a Joint Commitment. *Child Development, 89*(5), 1691–1703. <https://doi.org/10.1111/cdev.12816>
- Katz, E., Blumer, J. G., & Gurevitch, M. (1974). Utilization of mass communication by the individual. In J. G. Blumer & E. Katz (Eds.), *The uses of mass communications: Current perspectives on gratifications research* (pp. 19–34). SAGE.
- Kelley, H. (1967). Attribution Theory in Social Psychology. *Nebraska Symposium on Motivation, 15*, 192–238.
- 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>
- Kenworthy, J. B., & Miller, N. (2002). Attributional biases about the origins of attitudes: Externality, emotionality, and rationality. *Journal of Personality and Social Psychology, 82*(5), 693–707. <https://doi.org/10.1037/0022-3514.82.5.693>
- 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>
- 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>
- Kim, Y. (2015). Does Disagreement Mitigate Polarization? How Selective Exposure and Disagreement Affect Political Polarization. *Journalism & Mass Communication Quarterly, 92*(4), 915–937. <https://doi.org/10.1177/1077699015596328>
- Kim, Y., Hsu, S. H., & de Zúñiga, H. G. (2013). Influence of social media use on discussion

- network heterogeneity and civic engagement: The moderating role of personality traits. *Journal of Communication*, 63(3), 498–516. <https://doi.org/10.1111/jcom.12034>
- Koban, K., Stein, J., Eckhardt, V., & Ohler, P. (2018). Quid pro quo in Web 2.0. Connecting personality traits and Facebook usage intensity to uncivil commenting intentions in public online discussions. *Computers in Human Behavior*, 79, 9–18. <https://doi.org/10.1016/j.chb.2017.10.015>
- Kramer, R. M., Hanna, B. A., Su, S., & Wei, J. (2001). Collective identity, collective trust, and social capital: Linking group identification and group cooperation. In M. E. Turner (Ed.), *Groups at work: Theory and research* (pp. 173–196). Lawrence Erlbaum Associates.
- Kümpel, A. S., Karnowski, V., & Keyling, T. (2015). News Sharing in Social Media: A Review of Current Research on News Sharing Users, Content, and Networks. *Social Media + Society*, 1(2). <https://doi.org/10.1177/2056305115610141>
- Kwon, K. H., & Gruzd, A. (2017). Is offensive commenting contagious online? Examining public vs interpersonal swearing in response to Donald Trump’s YouTube campaign videos. *Internet Research*, 27(4), 991–1010. <https://doi.org/10.1108/IntR-02-2017-0072>
- Lea, M., O’Shea, T., Fung, P., & Spears, R. (1992). “Flaming” in computer-mediated communication: Observations, explanations, implications. In M. Lea (Ed.), *Contexts of computer-mediated communication* (pp. 89–112). Harvester-Wheatsheaf.
- Lea, M., & Spears, R. (1991). Computer-mediated communication, de-individuation and group decision-making. *International Journal of Man-Machine Studies*, 34(2), 283–301. [https://doi.org/10.1016/0020-7373\(91\)90045-9](https://doi.org/10.1016/0020-7373(91)90045-9)
- Li, C., Sun, Y., Ho, M. Y., You, J., Shaver, P. R., & Wang, Z. (2016). State narcissism and aggression: The mediating roles of anger and hostile attributional bias. *Aggressive Behavior*, 42(4), 333–345. <https://doi.org/10.1002/ab.21629>
- Lopes, B., & Yu, H. (2017). Who do you troll and Why: An investigation into the relationship between the Dark Triad Personalities and online trolling behaviours towards popular and less popular Facebook profiles. *Computers in Human Behavior*, 77, 69–76. <https://doi.org/10.1016/j.chb.2017.08.036>
- Lu, Y., & Myrick, J. G. (2016). Cross-cutting exposure on facebook and political participation: Unraveling the effects of emotional responses and online incivility. *Journal of Media Psychology*, 28(3), 100–110. <https://doi.org/10.1027/1864-1105/a000203>
- Malesza, M. (2020). The effects of the Dark Triad traits in prisoner’s dilemma game. *Current*

- Psychology*, 39(3), 1055–1062. <https://doi.org/10.1007/s12144-018-9823-9>
- Malik, A., Dhir, A., & Nieminen, M. (2016). Uses and Gratifications of digital photo sharing on Facebook. *Telematics and Informatics*, 33(1), 129–138. <https://doi.org/10.1016/j.tele.2015.06.009>
- Malle, B. F. (1999). How People Explain Behaviour: A New Theoretical Framework. *Personality and Social Psychology Review*, 3(1), 23–48. [https://doi.org/10.1207%2Fs15327957pspr0301\\_2](https://doi.org/10.1207%2Fs15327957pspr0301_2)
- Malle, B. F. (2011). Time to give up the dogmas of attribution. An alternative theory of behavior explanation. In *Advances in Experimental Social Psychology* (1st ed., Vol. 44). Elsevier Inc. <https://doi.org/10.1016/B978-0-12-385522-0.00006-8>
- Malle, B. F., & Korman, J. (2013). Attribution Theory. In D. S. Dunn (Ed.), *Oxford Bibliographies in Psychology* (p. No pagination). Oxford University Press.
- Marcus, G. E., MacKuen, M., Wolak, J., & Keele, L. (2006). The Measure and Mismeasure of Emotion. In D. P. Redlawsk (Ed.), *Feeling politics: Emotion in political information processing* (pp. 31–45). Palgrave Macmillan. <https://doi.org/10.1017/CBO9781107415324.004>
- Marcus, G. E., Neuman, W. R., & MacKuen, M. (2000). *Affective Intelligence and Political Judgment*. University of Chicago Press.
- Masullo, G. M., Lu, S., & Fadnis, D. (2021). Does online incivility cancel out the spiral of silence? A moderated mediation model of willingness to speak out. *New Media & Society*, 23(11), 3391–3414. <https://doi.org/10.1177/1461444820954194>
- McCrae, R. R., & Costa, P. T. (2003). *Personality in adulthood: A five-factor theory perspective*. Guilford.
- Moor, L., & Anderson, J. R. (2019). A systematic literature review of the relationship between dark personality traits and antisocial online behaviours. *Personality and Individual Differences*, 144, 40–55. <https://doi.org/10.1016/j.paid.2019.02.027>
- Moor, P. J., Heuvelman, A., & Verleur, R. (2010). Flaming on YouTube. *Computers in Human Behavior*, 26(6), 1536–1546. <https://doi.org/10.1016/j.chb.2010.05.023>
- Mouffe, C. (2000). *The Democratic paradox*. Verso.
- Muddiman, A. (2017). Personal and public levels of political incivility. *International Journal of Communication*, 11, 3182–3202.
- Muddiman, A., Pond-Cobb, J., & Matson, J. E. (2020). Negativity Bias or Backlash: Interaction With Civil and Uncivil Online Political News Content. *Communication Research*, 47(6), 815–837. <https://doi.org/10.1177/0093650216685625>

- 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>
- Ott, B. L. (2017). The age of Twitter: Donald J. Trump and the politics of debasement. *Critical Studies in Media Communication*, 34(1), 59–68.  
<https://doi.org/10.1080/15295036.2016.1266686>
- Oz, M., Zheng, P., & Chen, G. M. (2017). Twitter versus Facebook: Comparing incivility, impoliteness, and deliberative attributes. *New Media and Society*, 20(9), 3400–3419.  
<https://doi.org/10.1177/1461444817749516>
- Pabian, S., De Backer, C. J. S., & Vandebosch, H. (2015). Dark Triad personality traits and adolescent cyber-aggression. *Personality and Individual Differences*, 75, 41–46.  
<https://doi.org/10.1016/j.paid.2014.11.015>
- Palmgreen, P., & Rayburn, J. D. (1982). Gratifications sought and media exposure: An expectancy value model. *Communication Research*, 9(4), 561–580.  
<https://doi.org/10.1177/009365082009004004>
- Papacharissi, Z. (2002). The virtual sphere. *New Media & Society*, 4(1), 9–27.  
<https://doi.org/10.1177/14614440222226244>
- 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>
- Papacharissi, Z., & Rubin, A. M. (2000). Predictors of Internet Use. *Journal of Broadcasting & Electronic Media*, 44(2), 175–196. [https://doi.org/10.1207/s15506878jobem4402\\_2](https://doi.org/10.1207/s15506878jobem4402_2)
- Parks, L., & Guay, R. P. (2009). Personality, values, and motivation. *Personality and Individual Differences*, 47(7), 675–684. <https://doi.org/10.1016/j.paid.2009.06.002>
- Paulhus, D. L., & Williams, K. M. (2002). The Dark Triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality*, 36(6), 556–563. [https://doi.org/10.1016/S0092-6566\(02\)00505-6](https://doi.org/10.1016/S0092-6566(02)00505-6)
- Petit, J., & Carcioppolo, N. (2020). Associations between the Dark Triad and online communication behavior: A brief report of preliminary findings. *Communication Research Reports*, 37(5), 286–297. <https://doi.org/10.1080/08824096.2020.1862784>
- Phua, J., Jin, S. V., & Kim, J. (2017). Uses and gratifications of social networking sites for bridging and bonding social capital: A comparison of Facebook, Twitter, Instagram, and Snapchat. *Computers in Human Behavior*, 72, 115–122.  
<https://doi.org/10.1016/j.chb.2017.02.041>

- Postmes, T., Spears, R., & Lea, M. (1998). Breaching or Building Social Boundaries? *Communication Research*, 25(6), 689–715.  
<https://doi.org/10.1177/009365098025006006>
- Postmes, T., Spears, R., Sakhel, K., & de Groot, D. (2001). Social Influence in Computer-Mediated Communication: The Effects of Anonymity on Group Behavior. *Personality and Social Psychology Bulletin*, 27(10), 1243–1254.  
<https://doi.org/10.1177/01461672012710001>
- Protection, F. M. of J. and C. (2017). *Act to Improve Enforcement of the Law in Social Networks (Network Enforcement Act)*. German Federal Ministry of Justice and Consumer Protection.  
[https://www.bmju.de/SharedDocs/Gesetzgebungsverfahren/Dokumente/NetzDG\\_engl.pdf](https://www.bmju.de/SharedDocs/Gesetzgebungsverfahren/Dokumente/NetzDG_engl.pdf)
- Quandt, T. (2018). Dark Participation. *Media and Communication*, 6(4), 36–48.  
<https://doi.org/10.17645/mac.v6i4.1519>
- Rantanen, J., Metsäpelto, R. L., Feldt, T., Pulkkinen, L., & Kokko, K. (2007). Long-term stability in the Big Five personality traits in adulthood. *Scandinavian Journal of Psychology*, 48(6), 511–518. <https://doi.org/10.1111/j.1467-9450.2007.00609.x>
- Reeder, G. D. (2009). Mindreading: Judgments About Intentionality and Motives in Dispositional Inference. *Psychological Inquiry*, 20(1), 1–18.  
<https://doi.org/10.1080/10478400802615744>
- Reeder, G. D., & Brewer, M. B. (1979). A schematic model of dispositional attribution in interpersonal perception. *Psychological Review*, 86(1), 61–79.  
<https://doi.org/10.1037/0033-295X.86.1.61>
- Reeder, G. D., Kumar, S., Hesson-McInnis, M. S., & Trafimow, D. (2002). Inferences about the morality of an aggressor: The role of perceived motive. *Journal of Personality and Social Psychology*, 83(4), 789–803. <https://doi.org/10.1037/0022-3514.83.4.789>
- Reeder, G. D., Pryor, J. B., Wohl, M. J. A., & Griswell, M. L. (2005). On attributing negative motives to others who disagree with our opinions. *Personality and Social Psychology Bulletin*, 31(11), 1498–1510. <https://doi.org/10.1177/0146167205277093>
- Reeder, G. D., Vonk, R., Ronk, M. J., Ham, J., & Lawrence, M. (2004). Dispositional Attribution: Multiple Inferences About Motive-Related Traits. *Journal of Personality and Social Psychology*, 86(4), 530–544. <https://doi.org/10.1037/0022-3514.86.4.530>
- Rösner, L., & Krämer, N. C. (2016). Verbal Venting in the Social Web: Effects of Anonymity and Group Norms on Aggressive Language Use in Online Comments. *Social Media + Society*, 2(3). <https://doi.org/10.1177/2056305116664220>

- 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, L. (1977). The Intuitive Psychologist and His Shortcomings: Distortions in the Attribution Process. *Advances in experimental social psychology*, *10*, 173–220. [https://doi.org/10.1016/S0065-2601\(08\)60357-3](https://doi.org/10.1016/S0065-2601(08)60357-3)
- Rossini, P. (2019). Disentangling Uncivil and Intolerant Discourse. In R. G. Boatright, T. Shaffer, S. Sobieraj, & D. Young (Eds.), *A Crisis of Civility? Contemporary Research on Civility, Incivility, and Political Discourse*. (pp. 142–158). Routledge.
- Rossini, P. (2021). More Than Just Shouting? Distinguishing Interpersonal-Directed and Elite-Directed Incivility in Online Political Talk. *Social Media + Society*, *7*(2). <https://doi.org/10.1177/20563051211008827>
- Rossini, P. (2022). Beyond Incivility: Understanding Patterns of Uncivil and Intolerant Discourse in Online Political Talk. *Communication Research*, *49*(3), 399–425. <https://doi.org/10.1177/0093650220921314>
- Rost, K., Stahel, L., & Frey, B. S. (2016). Digital Social Norm Enforcement: Online Firestorms in Social Media. *PLoS ONE*, *11*(6), 1–26. <https://doi.org/10.1371/journal.pone.0155923>
- Rothmund, T., Gollwitzer, M., & Klimmt, C. (2011). Of virtual victims and victimized virtues: Differential effects of experienced aggression in video games on social cooperation. *Personality and Social Psychology Bulletin*, *37*(1), 107–119. <https://doi.org/10.1177/0146167210391103>
- Rowe, I. (2015). Civility 2.0: a comparative analysis of incivility in online political discussion. *Information Communication and Society*, *18*(2), 121–138. <https://doi.org/10.1080/1369118X.2014.940365>
- Ruggiero, T. E. (2018). Uses and Gratifications Theory in the 21st Century. *Refining Milestone Mass Communications Theories for the 21st Century*, *3*(1), 36–70. <https://doi.org/10.4324/9781315679402-4>
- Russo, S., & Amnå, E. (2016). The Personality Divide: Do Personality Traits Differentially Predict Online Political Engagement? *Social Science Computer Review*, *34*(3), 259–277. <https://doi.org/10.1177/0894439315582487>
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation. *American Psychologist*, *55*(1), 68–78. <https://doi.org/10.1037/0003066X.55.1.68>

- 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>
- Schäfer, M. S. (2011). Digital Public Sphere. In G. Mazzoleni, K. Barnhurst, K. Ikeda, M. Rousiley, & H. Wessler (Eds.), *The international encyclopedia of political communication* (pp. 322–328). Wiley-Blackwell.  
<https://doi.org/10.4018/9781599049472.ch246>
- Seidman, G. (2013). Self-presentation and belonging on Facebook: How personality influences social media use and motivations. *Personality and Individual Differences*, 54(3), 402–407. <https://doi.org/10.1016/j.paid.2012.10.009>
- Settle, J. E. (2018). *Frenemies: How social media polarizes America*. Cambridge University Press.
- Shmargad, Y., Coe, K., Kenski, K., & Rains, S. A. (2021). Social Norms and the Dynamics of Online Incivility. *Social Science Computer Review*, 089443932098552.  
<https://doi.org/10.1177/0894439320985527>
- Smock, A. D., Ellison, N. B., Lampe, C., & Wohn, D. Y. (2011). Facebook as a toolkit: A uses and gratification approach to unbundling feature use. *Computers in Human Behavior*, 27(6), 2322–2329. <https://doi.org/10.1016/j.chb.2011.07.011>
- Snyder, M., & Ickes, W. (1985). Personality and Social Behavior. In G. Lindzey & E. Aronson (Eds.), *Handbook of Social Psychology* (3rd ed., pp. 883–948). Addison-Wesley.
- 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>
- Soldz, S., & Vaillant, G. E. (1999). The Big Five Personality Traits and the Life Course: A 45-Year Longitudinal Study. *Journal of Research in Personality*, 33(2), 208–232.  
<https://doi.org/10.1006/jrpe.1999.2243>
- Soto, C. J. (2021). Do Links Between Personality and Life Outcomes Generalize? Testing the Robustness of Trait–Outcome Associations Across Gender, Age, Ethnicity, and Analytic Approaches. *Social Psychological and Personality Science*, 12(1), 118–130.  
<https://doi.org/10.1177/1948550619900572>
- 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>

- Stroud, N. J. (2010). Polarization and Partisan Selective Exposure. *Journal of Communication*, 60(3), 556–576. <https://doi.org/10.1111/j.1460-2466.2010.01497.x>
- 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 and Society*, 20(10), 3678-3699. <https://doi.org/10.1177/1461444818757205>
- Sundar, S. S., & Limperos, A. M. (2013). Uses and Grats 2.0: New Gratifications for New Media. *Journal of Broadcasting & Electronic Media*, 57(4), 504–525. <https://doi.org/10.1080/08838151.2013.845827>
- 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>
- Sydnor, E. (2019). *Disrespectful Democracy: The Psychology of Political Incivility*. Columbia University Press. <https://doi.org/10.7312/sydn18924>
- Thomas, D. R. (2006). A General Inductive Approach for Analyzing Qualitative Evaluation Data. *American Journal of Evaluation*, 27(2), 237–246. <https://doi.org/10.1177/1098214005283748>
- Tomasello, M. (2008). *Origins of human communication*. MIT Press.
- Tomasello, M. (2019). *Becoming human*. Harvard University Press.
- Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *The Behavioral and Brain Sciences*, 28(5), 675–691; discussion 691-735. <https://doi.org/10.1017/S0140525X05000129>
- Turel, O., & Qahri-Saremi, H. (2018). Explaining unplanned online media behaviors: Dual system theory models of impulsive use and swearing on social networking sites. *New Media & Society*, 20(8), 3050-3067. <https://doi.org/10.1177/1461444817740755>
- Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Margaret, S. (1987). Rediscovering the Social Group: A Self-Categorization Theory. In J. C. Turner, M. A. Hogg, P. J. Oakes, S. S. Reicher, & M. S. Wetherell (Eds.), *Rediscovering the Social Group: A Self-Categorization Theory*. Basil Blackwell.
- Valenzuela, S., Correa, T., & Gil de Zúñiga, H. (2018). Ties, Likes, and Tweets: Using Strong and Weak Ties to Explain Differences in Protest Participation Across Facebook and

- Twitter Use. *Political Communication*, 35(1), 117–134.  
<https://doi.org/10.1080/10584609.2017.1334726>
- Valera-Ordaz, L. (2019). Liberal individualist, communitarian, or deliberative? Analyzing political discussion on Facebook based on three notions of democracy. *International Journal of Communication*, 13, 1056–1076.
- Valeriani, A., & Vaccari, C. (2016). Accidental exposure to politics on social media as online participation equalizer in Germany, Italy, and the United Kingdom. *New Media and Society*, 18(9), 1857–1874. <https://doi.org/10.1177/1461444815616223>
- Van Duyn, E., & Muddiman, A. (2022). Predicting perceptions of incivility across 20 news comment sections. *Journalism*, 23(1), 134–152.  
<https://doi.org/10.1177/1464884920907779>
- van Geel, M., Goemans, A., Toprak, F., & Vedder, P. (2017). Which personality traits are related to traditional bullying and cyberbullying? A study with the Big Five, Dark Triad and sadism. *Personality and Individual Differences*, 106, 231–235.  
<https://doi.org/10.1016/j.paid.2016.10.063>
- Vogels, E. A. (2021). *The State of Online Harassment*. Pew Research Center.  
<https://www.pewresearch.org/internet/2021/01/13/personal-experiences-with-online-harassment/>
- 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>
- Weiner, B. (2000). Intrapersonal and interpersonal theories of motivation from an attributional perspective. In F. Salili, C. Y. Chiu, & Y. Y. Hong (Eds.), *Educational Psychology Review* (Vol. 12, Issue 1, pp. 1–14). <http://www.jstor.org/stable/23363504>
- Ziegele, M., Breiner, T., & Quiring, O. (2014). What Creates Interactivity in Online News Discussions? An Exploratory Analysis of Discussion Factors in User Comments on News Items. *Journal of Communication*, 64(6), 1111–1138.  
<https://doi.org/10.1111/jcom.12123>
- 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>

## **VI Research articles**

In the following, the four research articles of the cumulus are presented. The order of the articles corresponds to the order of the presented studies in the cumulus. Study I has been published in conference proceedings, and Study III has been published in an international journal. Studies II and IV have been submitted to international journals. Study II is currently under review, and Study IV is in revision and is permitted for resubmission.

# Article I

© 2020 Copyright held by the owner/author(s). Publication rights licensed to ACM. This is the author's version of the work. It is posted here for personal use only and not for redistribution. The definitive Version of Record was published in:

Kluck, J. P. & Krämer, N. C. (2020). It's the aggression, stupid! An examination of commenters' motives for using incivility in online political discussions. In *Proceedings of the 11th International Conference on Social Media and Society*, 164–173. <https://doi.org/10.1145/3400806.3400826>

# It's the aggression, stupid!

An examination of commenters' motives for using incivility in online political discussions.

Jan P. Kluck

Department of Social Psychology: Media and Communication, University of Duisburg-Essen, Germany, jan.kluck@uni-due.de

Nicole C. Krämer

Department of Social Psychology: Media and Communication, University of Duisburg-Essen, Germany, nicole.kraemer@uni-due.de

## ABSTRACT

Incivility in online political discussions is an urgent problem for society as it can have detrimental effects on democratic discourse. However, knowledge about people's inherent motives for behaving in an uncivil way is scarce. To close this gap, this online study employed a mixed-methods approach and surveyed 115 discussion commenters to explore their motives for communicating in an uncivil manner. Qualitative analyses revealed that in individual cases, people's motives for uncivil commenting can be very complex, and differ between distinct forms of incivility. Participants also indicated that they often try to counteract other comments. In contrast, quantitative data revealed that motives related to aggression were the key drivers for frequent uncivil commenting behavior among different forms of incivility. In sum, the study demonstrated that although aggressive motives are not the only predictors of uncivil commenting, they still seem to be a key factor for an individual's tendency to write uncivil comments.

## CCS CONCEPTS

•Human-centered computing~Human computer interaction (HCI)~Empirical studies in HCI

## KEYWORDS

Incivility, online political discussions, motivation, user comments

## 1 Introduction

Scholarship has noted that online communication has become increasingly uncivil, and that this tendency may decrease the Internet's potential value in fostering political discourse [6]. Indeed, incivility has been associated with various negative effects that might corrupt the formation of civic opinions. For instance, incivility has the potential to polarize attitudes [1] and can lead to less open-mindedness [3]. Moreover, it can discourage users from participating in online discussions [12], thus lowering the potential of the Internet to promote diversity in political debates.

Although the issue has garnered academic interest, the causes of uncivil behavior remain difficult to determine. However, the dissemination of incivility can only be counteracted if we understand the mechanisms that drive persons to act in uncivil ways. Various factors have been identified as fostering the prevalence of uncivil comments [e.g., 6, 11, 24], but less is known about users' underlying motivation for writing uncivil comments. One major problem is that incivility is a multidimensional concept, which encompasses a range of distinct antisocial communication behaviors [30]. However, research has largely neglected to differentiate between various forms of incivility when exploring which factors foster its occurrence. Since people's underlying motivation determines their behavior [8], identifying individuals' motivation for being uncivil might be key to understanding the dynamics of incivility in online discussions. To address this, we employed a unique approach by conducting a survey that incorporated qualitative and quantitative measures to interview 115 commenters of online political discussions. In this way, we sought to identify the motives people have when commenting in uncivil ways and how these motives differ between different forms of incivility. Moreover, we aimed to explore how individuals' personalities influence their motivation for uncivil commenting. To gain an understanding of the range of motives people may have when commenting uncivilly, we used qualitative measures to ask about specific situations in which commenters behaved uncivilly. Furthermore, quantitative data were collected to understand how people's general motives for commenting in online discussions influence their use of incivility. Additionally, the *Dark Triad* traits were assessed to examine whether they influence people's propensity for aggressive motives.

### 1.1 Incivility and its causes

Incivility has been defined as “features of discussion that convey an unnecessarily disrespectful tone toward the discussion forum, its participants, or its topics” [6]. However, Papacharissi [21] claimed that incivility differs from impoliteness, and should be regarded as “disrespect for the collective traditions of democracy” (p. 267), operationalized as the usage of antagonistic stereotypes, the verbalization of threats to democracy or the threat to other personal rights. Addressing this discrepancy, Stryker and colleagues [30] found that incivility is a multidimensional concept, and identified three dimensions on which incivility can function: utterance incivility (e.g., name-calling or vulgarity), discursive incivility

(e.g., preventing discussion or refusing to listen), and deception (e.g., exaggerating or misinforming others). However, it was proposed that people also use more subtle forms of attacking others in discussions, such as irony [2] or disrupting a discussion by posting comments that do not match the topic of discussion [29]. Although these forms are not directly considered to be uncivil, they appear to have the potential to lower the quality of a discussion and might thus be perceived as uncivil at least in some contexts.

Ultimately, however, the assessment of incivility is highly subjective. A discussion comment can be perceived as uncivil even when it was not meant to be, and conversely, a comment might be perceived as civil even when it was meant to be harmful. Therefore, it has been argued that the consequences of uncivil comments might be a better indicator for the assessment of incivility than the intent of the sender [4]. As such, we define incivility as *a discussion feature used by discussion participants, which intentionally or unintentionally threatens the discussion's potential for adding value to democratic discourse*. As incivility can manifest in many forms, by using this definition we not only investigate motives for commenting in a disrespectful tone, but also explore people's motives when they are deceiving others, violating democratic norms, and engaging in more subtle forms of incivility (e.g., irony and topic deviation).

Since uncivil behavior has become a serious problem, it is important to understand the drivers that induce people to comment in uncivil ways, as only then can effective interventions be developed. However, scholars have mainly focused on contextual factors to explain what causes uncivil behavior. For instance, people often lose their inhibitions when operating online due to the visual anonymity they experience; this has long been considered to be a main driver for uncivil online behavior [26]. Other research suggests that previous incivility might be a strong predictor of future uncivil online behavior [11, 24]. According to Gervais [11], emotional uncivil reactions to like-minded incivility might simply be induced by conformity processes (i.e., accepting and reusing like-minded messages), whereas offending uncivil reactions are triggered by uncivil disagreement because the resulting anger motivates recipients to “fight” (p. 400).

Indeed, people's willingness to perform certain behaviors depends on their goals and the expected outcomes [8]. Against this background, the *uses and gratifications approach* has often been employed to explain why people use certain online media and related features. Although this approach is not a coherent theory, it provides theoretical assumptions that help to investigate media usage behavior. According to this approach, people behave in a goal-directed and active manner when using media, have certain expectations for media use, and select specific media to fulfill their social and psychological needs [25]. Based on this idea, scholars have already determined distinct motives people have for writing comments. For instance, Springer, Engelmann, and Pfaffinger [28] identified cognitive motives (e.g., knowledge gain) and interactive motives (e.g., responding to journalists) as key drivers for commenting on online news. Similarly, other research found that people's motives for commenting online are related to information behavior (e.g., adding/correcting information), personal identity (e.g., expressing opinions), entertainment, and social interaction (e.g., debating) [9]. Further studies that explicitly focused on uncivil communication indicate that people often write uncivil comments to express anger, to react to perceived offenses, or simply to express critique or disagreement [19]. Research addressing more extreme forms of antisocial online behavior like trolling or cyberbullying found that people's motives can also predominantly aim to cause damage or to feel superior when commenting online [16, 27]. One would intuitively expect that – equivalent to other antisocial behavior – a person acting uncivilly in online discussion must also be pursuing an aggressive goal (e.g., causing harm or disrupting a discussion). On the other hand, incivility differs from antinormative behavior such as trolling and cyberbullying insofar as its goal is not necessarily to cause any harm or to provoke others. People might also use incivility to pursue other goals like interactive or cognitive aims. Since there is no evidence regarding which goals people pursue when commenting uncivilly, we ask (RQ1): *What goals do people have when commenting in an uncivil way in online political discussions?*

Whereas the aforementioned goals function in a rather direct way, people also have desires on a higher-order level that can affect their more immediate goals. For instance, it has been proposed that the experienced value of rewards resulting from social interaction can influence individuals' goals, and in turn predict either typical or dysfunctional social behavior [10]. In an analysis of different categories of social rewards, it was found that negative social potency, the subjective value experienced from “being cruel, callous and using others for personal gains”, is related to the malicious side of an individual's personality [10]. Moreover, it has been demonstrated that negative social potency can predict antisocial online behavior [7]. However, whereas Craker and March [7] conceptualized the value of social rewards as an immediate motivational force, we consider the experienced value of such rewards to be a precursor: the experienced reward value is likely to predict an individual's goals [10], which can be considered as the more immediate motivational force [8]. Therefore, we hypothesized that (H1) *the experienced value of negative social potency positively influences participants' frequency of uncivil commenting through the pathway of attributed motives*.

However, given the diverse forms of incivility, it seems unlikely that people have the same motives across all forms of incivility. For instance, the motives of people who deceive others (e.g., by spreading misleading information) probably differ substantially from the motives of people who directly insult others. To date there is hardly any evidence regarding how people's motives differ between distinct forms of incivility. To fill this gap, we ask (RQ2): *How do people's goals when commenting in an uncivil way differ between distinct forms of incivility?*

## 1.2 The impact of an individual's dark side

Although understanding people's motives for being uncivil might be key to understanding the causes thereof, the effect of an individual's personality on motivational processes should not be disregarded [22]. In this respect, antisocial online behavior has often been associated with the “darker side of personality”, illustrated by the Dark Triad personality model [18],

which includes the malicious but subclinical traits of narcissism, psychopathy, and Machiavellianism. Narcissism describes a person's tendency to feel grandiose, dominant and superior, and to be demanding. Subclinical psychopathy is characterized by high impulsivity and thrill-seeking but also by low empathy and anxiety, and people high in Machiavellianism have an increased tendency to manipulate others for their own personal gain [23]. Although these traits are very different, they are all offensive in particular ways [23] and all three have been found to correlate with antisocial behavior such as cyber-aggression, trolling behavior, or cyberbullying [18]. However, in the case of sexting (here, meaning disseminating images with explicit sexual content), it was demonstrated that when controlling for the motivation to perform such behavior, the Dark Triad traits were no longer independent predictors of non-consensual sext dissemination [5]. This indicates that the Dark Triad might not directly foster antinormative online behavior, but rather predict behavioral motives that encourage such behavior. Therefore, an examination of which individuals have predispositions for motives that foster uncivil online behavior may help to refine the understanding of what causes uncivil behavior. People scoring high on the Dark Triad traits, for instance, presumably have rather offensive motives for commenting in online discussions. Thus, it is assumed that *(H2) the Dark Triad traits positively affect participants' frequency of uncivil commenting through the pathway of motives that are primarily associated with antisocial online behavior.*

## 2 Methods

To explore which motives people have for behaving in uncivil ways, we employed a mixed-methods approach by collecting qualitative and quantitative data. Since research on individuals' motivation for uncivil commenting is scarce, this approach provides a more comprehensive perspective of what drives people to behave in such a manner. Given that online discussions can follow many different courses, we used a scenario-based approach, including qualitative questions about individuals' motives when posting a concrete uncivil comment in order to gain a better understanding of the situational goals people pursue when behaving uncivilly. People will also have specific traits such as psychological and social needs that guide their behavior in online discussions; therefore, quantitative questions were also included, targeting more general commenting behavior to elucidate the underlying discussion needs and motives that can foster uncivil commenting. The study was conducted as an online survey, and data were only collected from individuals who indicated that they comment in online political discussions. All analyses were conducted using *IBM SPSS 25* and *Amos 25*. The study was approved by the local ethics committee.

### 2.1 Sample

Participants from a German-speaking convenience sample were recruited mainly via Facebook groups related to political discussions and news comment sections, but also through online discussion forums and other social media platforms. Gift cards from an online retail store were offered as an incentive to take part in the study. Despite the recruitment approach and the incentive on offer, it proved difficult to acquire a sample of participants who actively comment (as only a small minority of Internet users post comments [31]). For the sake of validity of the results, however, we took great care to only include participants who actively comment. In sum, 152 individuals completed the survey. However, based on unrealistic completion time and/or implausible response behavior, 37 participants were excluded from the analyses, resulting in a total sample of 115 participants (71 females, 42 males, and 2 participants who did not specify their sex) aged between 19 and 70 years ( $M = 41.42$ ,  $SD = 14.19$ ). Most of the participants reported having a university degree (53%) or higher-track school-leaving qualifications (19%) as their highest level of education, and the majority were employees (42%) or university students (26%). Participants' political stance was mainly left-leaning ( $M = 31.20$ ,  $SD = 20.29$ ; from 0 = *far left* to 100 = *far right*). Concerning online discussion behavior, participants indicated that they commented in online political discussions on average 35 times a month ( $SD = 52.34$ ) and read comments of other users on average 83 times a month ( $SD = 76.77$ ).

### 2.2 Design and Procedure

To examine why people comment uncivilly, survey participants were presented with twelve different texts in random order. Each text was introduced as an example of a common communication behavior which is often found in online discussions, although ten of the texts described a specific type of incivility. If participants indicated that they had already used the respective type of incivility in a discussion comment, they were asked several open-ended questions about a specific situation in which they had commented in this way. Moreover, for further exploratory purposes, participants answered some scale-based questions regarding the perception of their own comment (which were not relevant for the present research question and manuscript). Subsequently, quantitative data were collected. Participants rated the perceived value of different social rewards and of the Dark Triad personality traits. Furthermore, they indicated their motives for commenting in online political discussions in general. At the end, participants answered questions regarding their demographics and their general online behavior.

### 2.3 Materials

We generated different descriptions for ten distinct forms of incivility; this selection was oriented to the literature [6, 21, 30] and sought to capture different dimensions of incivility. Forms of incivility belonging to deception were as follows: *disinformation* (deliberately sharing information that is not entirely true) and *exaggeration* (using exaggerated wording which

can cause misperceptions), as a rhetorical form of deception. We also included forms of incivility that represent ad-hominem attacks and disrespect: *insulting* (deliberately using language that is potentially perceived as insulting by others), *vulgarity* (using swear words) and *mockery* (making fun of other people's statements). Additionally, we addressed forms of incivility that violate democratic norms: *antagonistic stereotyping* (disliking others based on belonging to a certain social group), *voicing threats* (counteracting opposing groups with violence) and *denying opinion expression* (demanding that others should not express their opinion). We likewise included *irony/sarcasm* (making an assertion but simultaneously underlining that this is not one's true opinion) and *topic deviation* (deviating from the focus of discussion), because people often use more subtle forms of attacking others in discussions or of disrupting discussions [2, 29]. Notably, to prevent social desirability bias, the descriptions did not explicitly indicate that the respective communication behavior refers to incivility but rather implied that the respective behavior is common. For example, in case of vulgarity, the description stated: "In online political discussions, it happens again and again that people use swear words which other people might consider as inappropriate (e.g., 'fucking hell!')."¹ Additionally, two other ways of communicating in discussions (i.e., moderating between different opinions and pointing out false information) were included in the survey to reduce perceived negativity of the behaviors; however, data associated with these were not analyzed.

## 2.4 Measures

**Frequency of commenting and reading uncivil comments.** For each form of incivility, participants were asked how often they have read such comments and how often they have commented in such a fashion themselves. Both questions were rated on a 5-point Likert scale (from 1 = *never* to 5 = *very often*). To create a sum score of the frequency of uncivil commenting, we aimed to explore subscales among the different types of incivility. Although an exploratory factor analysis (EFA) with principal component analysis implied a three-factor solution, Horn's parallel analysis [14] indicated a single factor. Therefore, the final scale included all types of incivility, providing acceptable reliability (Cronbach's  $\alpha = .76$ ).

**Motives for uncivil commenting (qualitative exploration of motives).** If participants indicated having already commented in a respective uncivil way, they were instructed to think of a specific situation in which they had done so. To help participants to recall respective situations, three open-ended questions were asked (e.g., "On which political issue did you write the comment?"). To explore participants' motives for commenting in a specific uncivil way, they were asked to indicate their motives in a further open-ended question: "What goal were you pursuing by commenting in this way, i.e., what did you want to achieve by writing the comment?" For this question, we conducted a content analysis. Since comprehensive theoretical assumptions about individuals' motivation for uncivil commenting are lacking, we analyzed participants' answers in an inductive manner. This approach enables higher-order categories and theoretical assumptions to be derived from qualitative data [32]. As a result, 19 motive categories were derived. All key elements of the categories were described in a codebook, through which answers were coded with 1 or 0, depending on whether or not a motive category applied. Since answers were open-ended, multiple answers were possible. Two additional independent raters who were blind to the research questions coded the answers showing high overall agreement (from 81% to 99%). In case of dissent, motives were coded according to the majority. Subsequently, the distinct motive categories were summarized to six superordinate motive dimensions: information related motives (i.e., motives that focus on the mere distribution or correction of facts), discursive motives (i.e., motives that focus on the discourse with other), identity motives (i.e., motives that focus on demonstrating one's own social or political identity), aggressive motives (i.e., motives that focus on harming others), counteractive motives (i.e., motives that focus on counteracting inappropriate discussion behavior), affective motives (i.e., motives that focus on the direction of one's own emotions). See [Table 1](#) for details.

---

¹ All descriptions can be found in an online Appendix on the OSF website: <https://osf.io/ybj2s/>

**Table 1: Derived motive categories**

Motive categories	Examples	Motive descriptions (intercoder agreement in percentage)
Information-related motives	“Exposing the putative claim of responsibility as fake”	Correcting information and/or expose information as wrong (93%); adding and/or explaining facts in the discussion (88%)
Discursive motives	“That the person being addressed questions him/herself”	Increasing awareness of one's own comment or an issue addressed in the comment (92%); exchanging opinions and/or experiences, chatting with or answering questions of other discussion participants (92%); encouraging discussion participants to think about and/or rethink certain issues (89%)
Identity motives	“Valuing those who have been offended, rejecting offenders”	Signaling others that oneself is completely different from them (92%); signaling others that one feels connected to them (98%); indicating one's own explicit (political) position (87%)
Aggressive motives	“Annoying the author as much as possible”	Showing others that they are intellectually inferior (98%); annoying and/or provoking others as well as fueling conflicts (95%); embarrassing, mocking, and/or discrediting individuals or groups (92%)
Counteractive motives	“Putting hate in the discussion into perspective”	Making the discussion more factual, unemotional and/or reasonable (93%); Criticizing others by indicating that they behave improperly, have a reprehensible (political) point of view, or state inadequate arguments (81%); paying back other's negative behavior (97%)
Affective motives	“Getting rid of my frustration”	Expressing emotions and/or processing emotions (96%); goals that aim to regulate oneself in the communicative situation but also to reflect upon oneself in the situation (9%); amusing oneself or others and having fun (99%)

**General motives for commenting (quantitative exploration of motives).** To measure participants' motives for commenting in online political discussions in general, 25 items were adapted from Springer et al. [28], measuring motives for writing comments. We also adapted and modified 25 additional items from the *Motivation Aggression Questionnaire* [20] to capture aggressive motives for commenting. All items were rated on a 7-point Likert scale (from 1 = *does not apply at all* to 7 = *applies completely*). An EFA with principal component analysis and varimax rotation suggested a five-factor solution, which was confirmed by Horn's parallel analysis [14]. With a first series of EFAs with principal axis analysis and promax rotation for a fixed number of five factors, 25 items were removed based on low factor loadings (main factor < 0.50, other factors > 0.30). In this process, one factor was eliminated completely. Therefore, an additional EFA with a fixed number of four factors was computed, revealing the final structure<sup>2</sup>. The first factor, *aggression*, includes eight items targeting hostile intent (e.g., “to annoy others in the discussion”, Cronbach's  $\alpha = .92$ ); the second factor includes eight items comprising motives associated with knowledge gain, and was termed *cognition* (e.g., “to broaden my knowledge”, Cronbach's  $\alpha = .90$ ); the third factor contains six items which represent goals focusing on disagreeing, counteracting other comments, or expressing one's opinion, and was therefore labeled *positioning* (e.g., “to counteract inappropriate comments” but also “to express my opinion”, Cronbach's  $\alpha = .81$ ); and the fourth factor, *affect*, includes three items addressing individuals' feelings of anxiety and sadness (e.g., “when the subject of the discussion frightens me”, Cronbach's  $\alpha = .76$ ).

**Negative social potency.** Negative social potency, the experienced reward value from being cruel and harmful to others, was rated on the *Social Reward Questionnaire* (SRQ) [10]. The subscale consists of five items (rated on a 7-point Likert scale from 1 = *does not apply at all* to 7 = *applies completely*) including statements like “I enjoy embarrassing others” or “I enjoy seeing others get hurt” (Cronbach's  $\alpha = .75$ ).

**Dark Triad.** The Dark Triad personality traits were measured using the German version of the Short Dark Triad personality measure [17], with items rated on 7-point Likert scales (from 1 = *does not apply at all* to 7 = *applies completely*). For each dimension of the Dark Triad, participants rated nine items. Machiavellianism was assessed with items such as “Most people can be manipulated” (Cronbach's  $\alpha = .82$ ), narcissism was assessed with items such as “People see me as a natural

<sup>2</sup> The items and their factor loadings can be found in an online Appendix on the OSF website: <https://osf.io/ybj2s/>

leader” (Cronbach’s  $\alpha = .74$ ), and psychopathy was assessed with items such as “People often say I’m out of control” (Cronbach’s  $\alpha = .68$ ).

### 3 Results

Descriptive analyses revealed that participants do indeed engage in uncivil commenting, albeit rather infrequently (Table 2). Whereas disinformation and voicing threats were employed with the lowest frequency, participants stated using irony/sarcasm most often.

**Table 2: Means and standard deviations of frequency participants stated to read and post uncivil comments in political online discussions.**

<i>N</i> = 115	Frequency of reading uncivil comment		Frequency of posting uncivil comment	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Disinformation	4.23	0.99	1.10	0.32
Exaggeration	4.31	0.97	1.46	0.87
Insulting	4.55	0.85	1.46	0.82
Vulgarity	4.18	1.01	1.29	0.57
Mockery	4.38	0.88	1.65	0.90
Antagonistic stereotyping	4.61	0.68	1.84	1.09
Voicing threats	3.91	1.14	1.15	0.57
Denying opinion expression	3.38	1.12	1.27	0.72
Topic deviation	4.45	0.81	1.42	0.81
Irony/sarcasm	4.44	0.83	2.04	1.17

Regarding participants’ motivation for uncivil commenting in online political discussions (*RQ1*) and how these motives differ between distinct forms of incivility (*RQ2*), qualitative analyses indicated that above all, counteractive motives seem to be an important main driver for insulting, mockery, and antagonistic stereotyping. Moreover, counteractive motives are predominant as drivers for topic deviation and opinion expression. For voicing threats, participants indicated most frequently having discursive and identity related motives for commenting in a respective way. For disinformation and exaggeration, informational motives were most frequently indicated and for using vulgarity, participants seem to have mainly affective motives. Moreover, information centric motives and counteractive motives were mostly cited as drivers for irony/sarcasm. Notably, aggressive motives were not present in any of the different types of incivility as the primary driver. However, for insulting and mockery, individuals still seem to have aggressive motives. Likewise, after affective motives, identity motives were indicated the most frequently for the usage of vulgarity. See Table 3 for an overview.

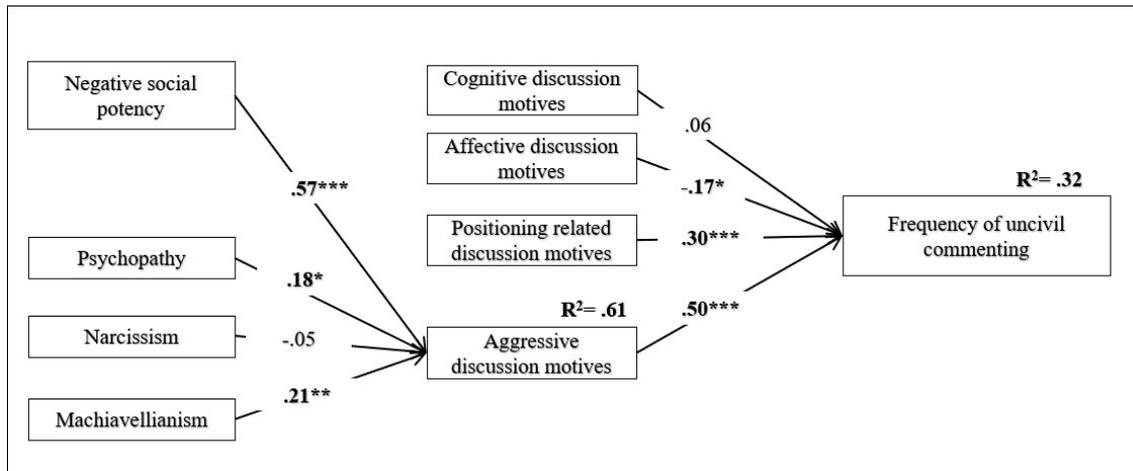
**Table 3: Proportion of motives indicated for the different forms of incivility.**

Form of Incivility	n	Motive categories					
		Information	Discourse	Identity	Aggression	Counter-action	Affect
Disinformation	10	n = 4 , 33.3%	n = 3, 25.0%	n = 0 0.0%	n = 1, 8.3%	n = 3, 25.0%	n = 1 8.3%
Exaggeration	31	n = 9, 25.0%	n = 8, 22.2%	n = 4, 11.1%	n = 5, 13.9%	n = 7, 19.4%	n = 3, 8.3%
Insulting	35	n = 5, 13.2%	n = 5, 13.2%	n = 5, 13.2%	n = 7, 18.4%	n = 11, 28.9%	n = 5, 13.2%
Vulgarity	27	n = 1, 3.8%	n = 0, 0.0%	n = 6, 23.1%	n = 4, 15.4%	n = 5, 19.2%	n = 10, 38.5%
Mockery	50	n = 8, 15.4%	n = 6, 11.5%	n = 3, 5.8%	n = 11, 21.2%	n = 19, 36.5%	n = 5, 9.6%
Antagonistic Stereotyping	54	n = 13, 23.2%	n = 7, 12.5%	n = 12, 21.4%	n = 4, 7.1%	n = 16, 28.6%	n = 4, 7.1%
Voicing threats	8	n = 1, 16.7%	n = 2, 33.3%	n = 2, 33.3%	n = 0, 0.0%	n = 1, 16.7%	n = 0, 0.0%
Denying opinion expression	20	n = 5, 22.7%	n = 2, 9.1%	n = 2, 9.1%	n = 1, 4.5%	n = 10, 45.5%	n = 2, 9.1%
Topic deviation	33	n = 9, 27.3%	n = 3, 9.1%	n = 4, 12.1%	n = 3, 9.1%	n = 13, 39.4%	n = 1, 3.0%
Irony/sarcasm	64	n = 12, 18.8%	n = 9, 14.1%	n = 9, 14.1%	n = 11, 17.2%	n = 12, 18.8%	n = 11, 17.2%

Note. Multiple answers were possible; some answers (n = 19) could not be allocated to a motive.

Examining quantitatively which motives might foster uncivil commenting (*RQ1*) and what influence negative social potency and the Dark Triad have on motives that are associated with antisocial behavior (*H1* & *H2*), we conducted a path analysis and maximum likelihood estimation. For the anticipated “antisocial” motives, we employed the aggression factor that was revealed by the EFA. Indirect effects were tested by using bias-corrected bootstrapping with 5000 resamples (95% confidence interval). The model fit well: [15]:  $\chi^2(17) = 18.425, p < .363, \chi^2/df = 1.08, CFI = 1.00, TLI = .99, RMSEA = .027$  (90% CI [.00, .09]), SRMR = .073. The model is visualized in Figure 1.

Whereas *cognitive motives* did not influence uncivil commenting, *positioning* and *aggressive motives* (as the strongest predictor) positively affected participants’ frequency of uncivil commenting. In contrast, *affect-related motives* negatively predicted the frequency of uncivil commenting. Regarding the experienced value of negative social potency (*H1*), the results revealed a significant positive influence on aggression-related motives. Likewise, the experienced value of negative social potency indirectly predicted the frequency of uncivil commenting ( $\beta = -.02, p < .555, 95\% \text{ CI } [-0.11, 0.06]$ ). Similarly, psychopathy and Machiavellianism significantly predicted aggressive motives. As expected, through this pathway, both traits indirectly affected the frequency of uncivil commenting (psychopathy:  $\beta = .09, p < .015, 95\% \text{ CI } [0.02, 0.17]$  and Machiavellianism:  $\beta = .42, p < .005, 95\% \text{ CI } [0.03, 0.21]$ ). However, narcissism did not influence participants’ predisposition for aggressive discussion motives and thus did not indirectly affect the frequency of uncivil commenting ( $\beta = -.02, p < .555, 95\% \text{ CI } [-0.11, 0.06]$ ). Nevertheless, *H2* was partially confirmed. Taken together, aggressive motives were the strongest predictor for the frequency of uncivil commenting behavior. Moreover, this tendency was affected by participants’ levels of psychopathy, Machiavellianism and their experienced value of negative social potency.



**Figure 1: Path model ( $N = 115$ ). Standardized effect sizes are reported; \* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$ . Model fit:  $\chi^2(17) = 18.425, p < .363, \chi^2/df = 1.08, CFI = 1.00, TLI = .99, RMSEA = .027$  (90% CI [.00, .09]), SRMR = .073.**

We also quantitatively examined whether the influence of distinct motives on uncivil commenting differs between the distinct forms of incivility ( $RQ2$ ). Therefore, for each type of incivility, we calculated a series of ordinary least square regressions that included the different motives as predictors and participants' indicated frequency of using the respective type of incivility as the criterion (see Table 4). As before, *aggression-related motives* significantly predicted the frequency of use among all forms of incivility, although the predictive power differed between the distinct types. In addition, *positioning* significantly predicted the use of irony/sarcasm, mockery, and antagonistic stereotyping. In the latter case, *positioning* was even a better predictor than aggression motives. Furthermore, results indicate that *cognition* and *affect-related motives* only predict the use of exaggeration. Whereas *cognitive motives* positively predicted the frequency of using exaggeration, affective motives were negatively related to the frequency of using exaggeration.

**Table 4: Regression analyses with different motive dimensions as predictors of frequency of using specific types of incivility.**

Frequency of using type of incivility $N = 115$	Predictors $\beta$				$R^2$
	Cognition	Affect	Positioning	Aggression	
Disinformation	.12	.15	-.12	.28**	.11
Exaggeration	.24*	-.23*	.10	.31**	.16
Insulting	.02	-.16	.12	.41***	.17
Vulgarity	-.06	-.10	.04	.41***	.19
Mockery	-.03	-.09	.25*	.27**	.10
Antagonistic stereotyping	-.07	-.12	.41***	.21*	.15
Voicing threats	-.01	-.09	.14	.31**	.10
Denying opinion expression	-.04	-.04	.16	.30**	.10
Topic deviation	.12	-.09	.13	.30**	.10
Irony/sarcasm	.06	-.10	.24*	.29**	.11

Note. Standardized regression coefficients are reported; \* $p < .05$ , \*\*  $p < .01$ , \*\*\* $p < .001$ .

## 4 Discussion

The current study was designed to obtain insights into people's motivation for acting uncivilly in online political discussions. Therefore, we inductively derived different motives from participants' qualitative answers. Qualitative results showed that counteractive motives (i.e., putting the discussion into perspective and rebuking others) were dominantly cited drivers for expressing insults, mockery, and antagonistic stereotypes, along with denying opinion expressions and deviating from the discussion topic. When purposefully misinforming others or using exaggerations, participants most frequently professed to having information-centered motives (i.e., correcting/adding information). Voicing threats was mostly linked to discourse-related motives (i.e., exchanging views or drawing attention) and identity motives (i.e., highlighting one's own stance), whereas expressions of vulgarity were mostly attributed to affective motives (i.e., expressing and regulating emotions). Importantly, aggressive motives (i.e., inducing conflict, provoking or derogating others) were not cited as main drivers for any of the identified behaviors. Although participants' answers were limited to isolated cases of uncivil

commenting, they provide valuable insight into understanding motivations for writing uncivil comments. Notwithstanding the participants self-reported drivers, path analyses of the quantitative data revealed that aggressive discussion motives (manifested in causing harm, unsettling others, and disrupting the discussion) are main drivers for frequent uncivil commenting. Regression analyses confirmed that these motives predict the frequency of uncivil commenting across all distinct types of incivility. In sum, results suggest that most people report rather unaggressive motives for specific uncivil comments – but findings also show that individuals who frequently behave uncivilly are mostly driven by aggressive discussion motives. Moreover, it was found that people's experienced value from negative social potency positively predicts the incidence of aggressive motives. Likewise, persons with higher scores on psychopathy and Machiavellianism were more likely to have such motives.

Qualitative data demonstrated that motives related to counteraction are important drivers for uncivil commenting. Although answers suggest that aggression-centered motives can provoke especially insults and mockery, participants mostly indicated that even through using these types of incivility, they were trying to counteract. Likewise, antagonistic stereotyping seems to be mainly motivated by counteractive goals. It is conceivable that in such cases, negative emotions like anger foster individuals' willingness to act uncivilly through the desire to fight back [11]. In contrast, topic deviation and denying opinion expressions, which also seem to be mainly motivated by counteractive goals, might be an avoidance-related way of coping with others' antinormative behavior. Interestingly, topic deviation was considered to be a more subtle form of negative online behavior [29], whereas denying opinion expression can be considered as a restriction of the individual's freedom of speech and therefore perceived as a more democracy-threatening behavior [21]. Thus, people might perform incivility in very different ways, while pursuing similar goals. Regarding topic deviation and antagonistic stereotyping, results further indicate that people often have information-centered motives. Such motives may be considered rather social because they put into perspective the public discourse [9]. In our case, these motives are probably related to counteractive motives insofar as people might inform others about negative facts regarding antagonistic social groups. That said, people frequently stated that they stereotype others based on identity-related motives. Thus, such motives might also be driven by negative emotions and therefore culminate in dysfunctional behavior like uncivil commenting. Participants' use of exaggerations likewise seems to be driven by information-centered motives, discursive motives, and counteractive motives. Although exaggeration is usually used to obscure the truth [30], it might therefore be the case that people are actually motivated to improve the discussion, such as by adding information or by expressing criticism. Furthermore, participants mainly stated having affect- or identity-related motives when commenting in vulgar ways. Thus, people might aim to highlight experienced emotions by expressing vulgarity instead of deliberately impairing the discussion. Another interesting finding is that information-related motives were mostly indicated for spreading disinformation in discussions. Therefore, some people might accept a loss of accuracy to exemplify or simplify information. Although the number of persons who reported a history of spreading disinformation or having voiced threats in an online discussion is relatively small, it remains remarkable that in a sample of 115 people, at least four (or three, respectively) people admitted having used these extreme forms. Nevertheless, the generalizability of the answers is limited. Finally, compared to the other forms of incivility, we did not identify a dominant pattern of motives for the use of irony/sarcasm in our data. Since irony/sarcasm was also proposed as a more subtle form of attacking others [2], people might use it as rhetorical act when pursuing different goals. However, although irony/sarcasm might be often employed to attack others, it is not necessarily used in an uncivil manner.

Overall, qualitative analyses indicated that, on a case-on-case basis, motivations for behaving uncivilly are complex and differ between distinct forms of incivility. Nevertheless, participants mostly indicated having counteractive motives and therefore probably use incivility strategically. Qualitative data referred only to individual cases of uncivil commenting to gain a better understanding of the range of motives which people have for acting uncivilly. Therefore, we also quantitatively analyzed individuals' motives for acting uncivilly and investigated how these motives affect people's tendency to write uncivil comments.

In contrast to the qualitative data, quantitative data demonstrated that aggressive motives are key drivers for frequent uncivil behavior. Motives related to (political) positioning also positively predicted the frequency, whereas affect-related motives negatively predicted the frequency of uncivil commenting. Previous studies have demonstrated that people have aggressive motives for habitual antisocial online behavior [e.g., 16, 27], and incivility appears to be no exception. Nevertheless, the present study provides evidence that other motives (i.e., positioning oneself by disagreeing, counteracting or expressing one's opinion) can facilitate the use of uncivil commenting too. Presumably, people who have the primary motivation of positioning themselves during online discussions prefer to participate in discussions in which they are faced with antagonistic attitudes. Since the confrontation with opposing attitudes is associated with increased anger and dislike [13], people who want to take a stance in discussions might be more likely to comment uncivilly because they have to cope with undesired emotions. In contrast, people with affective motives (i.e., feeling sad and/or anxious) when writing comments might be generally more conflict-avoidant and therefore less likely to comment in an uncivil way.

When examining the influence of different motives for each type of incivility separately (*RQ2*), results showed that aggressive motives positively predict the frequency of uncivil commenting across all different types. In contrast, positioning related motives only predicted people's frequency of antagonistic stereotyping and the use of irony/sarcasm. Thus, individuals might be more likely to counter a contradictory stance by disparaging antagonistic social groups, and irony/sarcasm might be the rhetorical device best suited to do so. Likewise, affect-related motives solely negatively predicted people's frequency of using exaggerations. To explain this, it might be speculated that when persons feel sad or anxious, they attempt to calm down the discussion and thus endeavor to avoid the use of exaggeration. Although cognitive motives did not

affect the overall score regarding the frequency of uncivil commenting, analyses on the individual types of incivility revealed that people who participate in online discussions to gain knowledge are more likely to use exaggerations. Again, it may be speculated that people use particular caution in keeping the discussion as factual as possible, but use exaggerations to elicit opinion expressions in an effort to better understand others' perspectives. Taken together, quantitative analyses revealed that aggressive motives are the most important predictor for frequent uncivil commenting, even when differentiating between distinct forms of incivility.

Notably, qualitative and quantitative data seem to contradict each other: overall, participants mostly cited counteractive motives for specific acts of incivility, but aggressive discussion motives were the strongest predictor for frequent uncivil commenting. It may be that people are particularly likely to recall incivility that they consider to have been justified, explaining why aggressive motives were not cited as dominant drivers. However, individuals who comment uncivilly are also more likely to have aggressive motives for joining online discussions.

We further hypothesized that individuals' experienced value from negative social potency positively influences the frequency of uncivil commenting through the pathway of discussion motives that are usually associated with antisocial online behavior (*H1*). Indeed, negative social potency positively affected the frequency of uncivil commenting through the pathway of aggressive motives. Therefore, on a higher-order level, when people have aggressive motives for commenting in online discussions, they generally expect to receive gratification from being cruel or causing harm to others. Thus, people who experience high value from negative social potency also aim to cause harm when commenting in online discussions and consequently write uncivil comments more frequently.

Lastly, as the Dark Triad has been found to foster people's antisocial online behavior [18], we argued that this tendency would be mediated by motives associated with antisocial behavior (*H2*). We found that psychopathy and Machiavellianism affected the frequency of uncivil commenting through the pathway of aggressive motives. Since people scoring high on these traits were also shown to be more prone to engage in antisocial online behavior like trolling [7], these individuals might be more likely to have aggressive motives because they behave in a more destructive and strategic manner. Notably, narcissism did not affect participants' aggressive motives. In fact, narcissism showed the weakest and most inconsistent relationship to distinct antisocial online behaviors, suggesting that narcissism is not as "dark" as psychopathy and Machiavellianism [18]. However, people who score high on narcissism possibly have other motives for commenting in online political discussions (e.g., motives focusing on feelings of superiority) that were not assessed by the current study.

## 4.1 Limitations

Some important limitations of the current study need to be acknowledged. First, the sample studied is relatively small and has limited representativeness. Therefore, the results should be generalized with caution. However, as it is well known that only a small proportion of Internet users actively posts comments, we successfully acquired a relevant sample of people who indicated engaging in online discussion. A further limitation lies in the self-report of participants' answers: although not explicitly indicated as incivility, the descriptions of the different forms of incivility were probably understood as antinormative behavior, potentially leading to a social desirability bias. Furthermore, we only investigated a limited selection of different types of incivility. As this study provides evidence that the particular manifestation of incivility employed by individuals matters, future research should consider further types of incivility. Finally, when examining the impact of people's general motives for commenting in online political discussions, we also provided a limited choice of possible motives, and it is possible that people have more diverse motives for discussing politics online.

## 4.2 Conclusions

We found that people can have very diverse motives for behaving uncivilly in online political discussions. Moreover, it was shown that motives can differ depending on the type of incivility that is used. The qualitative findings indicated that for specific discussion incidents, people often report writing uncivil comments to counteract experienced norm violations. This implies that people do not use incivility in a completely uncontrolled manner, but rather employ it strategically to change unwanted situations. On the other hand, the results also indicate that people are more likely to act uncivilly if they have aggressive discussion motives, irrespective of which form of incivility is employed. Therefore, future research should primarily concentrate on examining how aggressive motives can be reduced when designing interventions to combat incivility. Nevertheless, considering the complexity of motives for uncivil commenting as revealed by the qualitative data, future research should further investigate how other motives might affect people's tendency for uncivil commenting.

## ACKNOWLEDGMENTS

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

## REFERENCES

- [1] Ashley A. Anderson, Dominique Brossard, Dietram A. Scheufele, Michael A. Xenos, and Peter Ladwig. 2014. The "Nasty Effect:" Online Incivility and Risk Perceptions of Emerging Technologies. *Journal of Computer-Mediated Communication* 19, 3 (Apr. 2014), 373–387.
- [2] Ashley A. Anderson and Heidi E. Huntington. 2017. Social Media, Science, and Attack Discourse: How Twitter Discussions of Climate Change Use Sarcasm and Incivility. *Science Communication* 39, 5 (Oct. 2017), 598–620. DOI:<https://doi.org/10.1177/1075547017735113>.

- [3] Porismita Borah. 2014. Does It Matter Where You Read the News Story? Interaction of Incivility and News Frames in the Political Blogosphere. *Communication Research* 41, 6 (Aug. 2014), 809–827. DOI:<https://doi.org/10.1177/0093650212449353>.
- [4] Gina Masullo Chen, Ashley Muddiman, Tamar Wilner, Eli Parisser, and Natalie J. Stroud. 2019. We Should Not Get Rid of Incivility Online. *Social Media + Society* 5, 3 (Jul. 2019), 5 pages. DOI:<https://doi.org/10.1177/2056305119862641>.
- [5] Elizabeth M. Clancy, Bianca Klettke, and David J. Hallford. 2019. The dark side of sexting – Factors predicting the dissemination of sexts. *Computers in Human Behavior* 92 (Mar. 2019), 266–272. DOI:<https://doi.org/10.1016/j.chb.2018.11.023>.
- [6] Kevin Coe, Kate Kenski, and Stephen A. Rains. 2014. Online and Uncivil? Patterns and Determinants of Incivility in Newspaper Website Comments. *Journal of Communication* 64, 4 (Aug. 2014), 658–679. DOI:<https://doi.org/10.1111/jcom.12104>.
- [7] Naomi Craker and Evita March. 2016. The dark side of Facebook: The Dark Tetrad, negative social potency, and trolling behaviours. *Personality and Individual Differences* 102 (Nov. 2016), 79–84. DOI:<https://doi.org/10.1016/j.paid.2016.06.043>.
- [8] Edward L. Deci and Richard M. Ryan. 2000. The “What” and “Why” of Goal Pursuits: Human Needs and the Self-Determination of Behavior. *Psychological Inquiry* 11, 4 (Oct. 2000), 227–268. DOI:[https://doi.org/10.1207/S15327965PLI1104\\_01](https://doi.org/10.1207/S15327965PLI1104_01).
- [9] Nicholas Diakopoulos and Mor Naaman. 2011. Towards quality discourse in online news comments. In *Proceedings of the ACM 2011 conference on computer supported cooperative work (CSCW '11)*, March 19–23, 2011, Hangzhou, China. ACM Inc., New York, NY, 133–142. DOI:<https://doi.org/10.1145/1958824.1958844>.
- [10] Lucy Foulkes, Essi Viding, Eamon McCrory, and Craig S. Neumann. 2014. Social Reward Questionnaire (SRQ): Development and validation. *Frontiers in Psychology* 5 (Mar. 2014), 8 pages. DOI:<https://doi.org/10.3389/fpsyg.2014.00201>.
- [11] Bryan T. Gervais. 2017. More than Mimicry? The Role of Anger in Uncivil Reactions to Elite Political Incivility. *International Journal of Public Opinion Research* 29, 3 (Apr. 2017), 384–405. DOI:<https://doi.org/10.1093/ijpor/edw010>.
- [12] Soo-Hye Han and LeAnn M. Brazeal. 2015. Playing Nice: Modeling Civility in Online Political Discussions. *Communication Research Reports* 32, 1 (Jan. 2015), 20–28. DOI:<https://doi.org/10.1080/08824096.2014.989971>.
- [13] Ariel Hasell and Brian E. Weeks. 2016. Partisan Provocation: The Role of Partisan News Use and Emotional Responses in Political Information Sharing in Social Media. *Human Communication Research* 42, 4 (Oct. 2016), 641–661. DOI:<https://doi.org/10.1111/hcre.12092>.
- [14] John L. Horn. 1965. A rationale and test for the number of factors in factor analysis. *Psychometrika* 30 (Jun. 1965), 179–185. DOI:<https://doi.org/10.1007/BF02289447>.
- [15] Li-Tze Hu and Peter M. Bentler. 1999. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling* 6, 1 (Jan. 1999), 1–55. DOI:<https://doi.org/10.1080/10705519909540118>.
- [16] Andreas König, Mario Gollwitzer, and Georges Steffgen. 2010. Cyberbullying as an Act of Revenge? *Australian Journal of Guidance and Counselling*, 20, 2 (Dec. 2010), 210–224. DOI:<https://doi.org/10.1375/ajgc.20.2.210>.
- [17] Marta Malesza, Pawel Ostaszewski, Susanne Büchner, and Magdalena C. Kaczmarek. 2017. The Adaptation of the Short Dark Triad Personality Measure – Psychometric Properties of a German Sample. *Current Psychology*. (Aug. 2017), 10 pages. DOI:<https://doi.org/10.1007/s12144-017-9662-0>.
- [18] Lily Moor and Joel R. Anderson. 2019. A systematic literature review of the relationship between dark personality traits and antisocial online behaviours. *Personality and Individual Differences* 144 (Jul. 2019), 40–55. DOI:<https://doi.org/10.1016/j.paid.2019.02.027>.
- [19] Peter J. Moor, Ard Heuvelman, and Ria Verleur. 2010. Flaming on YouTube. *Computers in Human Behavior* 26, 6 (Nov. 2010), 1536–1546. DOI:<https://doi.org/10.1016/j.chb.2010.05.023>.
- [20] Ioan M. Ohlsson and Jane L. Ireland. 2011. Aggression and offence motivation in prisoners: Exploring the components of motivation in an adult male sample. *Aggressive Behavior* 37, 3 (Jan. 2011), 278–288. DOI:<https://doi.org/10.1002/ab.20386>.
- [21] Zizi Papacharissi. 2004. Democracy online: Civility, politeness, and the democratic potential of online political discussion groups. *New Media and Society* 6, 2 (Apr. 2004), 259–283. DOI:<https://doi.org/10.1177/1461444804041444>.
- [22] Laura Parks, and Russell P. Guay. 2009. Personality, values, and motivation. *Personality and Individual Differences* 47, 7 (Nov. 2009), 675–684. DOI:<https://doi.org/10.1016/j.paid.2009.06.002>.
- [23] Delroy L. Paulhus and Kevin M. Williams. 2002. The Dark Triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality* 36, 6 (Dec. 2002), 556–563. DOI:[https://doi.org/10.1016/S0092-6566\(02\)00505-6](https://doi.org/10.1016/S0092-6566(02)00505-6).
- [24] Leonie Rösner and Nicole C. Krämer. 2016. Verbal Venting in the Social Web: Effects of Anonymity and Group Norms on Aggressive Language Use in Online Comments. *Social Media + Society* 2, 3 (Sep. 2016), 13 pages. DOI:<https://doi.org/10.1177/2056305116664220>.
- [25] Thomas E. Ruggiero. 2000. Mass communication and society uses and gratifications theory in the 21st Century. *Mass communication & society* 3, 1 (Nov. 2000), 3–37. DOI:<https://doi.org/10.1207/S15327825MCS0301>.
- [26] Athur D. Santana. 2014. Virtuous or Vitriolic - The effect of anonymity on civility in online newspaper reader comment boards. *Journalism Practice* 8, 1 (Jan. 2014), 18–33. DOI:<https://doi.org/10.1080/17512786.2013.813194>.
- [27] Pnina Shachaf, and Noriko Hara. 2010. Beyond vandalism: Wikipedia trolls. *Journal of Information Science* 36, 3 (Apr. 2010), 357–370. DOI:<https://doi.org/10.1177/0165551510365390>. DOI:<https://doi.org/10.1080/10584609.2010.542360>.
- [28] Nina Springer, Ines Engelmänn, and Christian Pfaffinger. 2015. User comments: motives and inhibitors to write and read. *Information, Communication & Society* 18, 7 (Jul. 2015), 798–815. DOI:<https://doi.org/10.1080/1369118X.2014.997268>.
- [29] Natalie J. Stroud, Ashley Muddiman, and Joshua M. Scacco. 2015. Changing Deliberative Norms on News Organizations’ Facebook Sites. *Journal of Computer-Mediated Communication* 20, 2 (Mar. 2015), 188–203. DOI:<https://doi.org/10.1111/jcc4.12104>.
- [30] Robin Stryker, Bethany A. Conway, and J. Taylor Danielson. 2016. What is political incivility? *Communication Monographs* 83, 4 (Oct. 2016), 535–556. DOI:<https://doi.org/10.1080/03637751.2016.1201207>.
- [31] Na Sun, Patrick P. Rau, and Liang Ma. 2014. Understanding lurkers in online communities: A literature review. *Computers in Human Behavior* 38 (Sep. 2014), 110–117. DOI:<https://doi.org/10.1016/j.chb.2014.05.022>.
- [32] David R. Thomas. 2006. A General Inductive Approach for Analyzing Qualitative Evaluation Data. *American Journal of Evaluation* 27, 2 (Jun. 2006), 237–246. DOI:<https://doi.org/10.1177/1098214005283748>.

# Article II

## **Unwilling, intolerable views, or incapable? – How different types of incivility can be distinguished based on attributions about cooperative communication**

Submitted manuscript

Jan P. Kluck<sup>a,\*</sup>, Marike Bormann<sup>b</sup>, Alexander Rieß<sup>a</sup>, and Nicole C. Krämer<sup>a</sup>

<sup>a</sup> *Social Psychology: Media and Communication, University of Duisburg-Essen, Germany*

<sup>b</sup> *Department of Social Sciences, Heinrich Heine University Düsseldorf, Germany*

### **Acknowledgments:**

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

### **Declaration of interest:**

None

### **Compliance with Ethical Standard:**

The procedure performed in this study was approved by the ethics committee of the Department of Computer Science and Applied Cognitive Science of the University Duisburg-Essen and is in accordance with the standards of the institutional research committee and with the 1964 Helsinki Declaration, including later amendments or comparable ethical standards.

## **Unwilling, intolerable views, or incapable? – How different types of incivility can be distinguished based on attributions about cooperative communication**

### **Abstract**

Incivility in online discussions has gained notable attention from both academia and the public. However, incivility has become a fuzzy concept as definitions and operationalizations vary in the literature. We argue that this issue results from the lack of a common theoretical basis that allows categorizing different types of incivility. Therefore, we draw on assumptions about cooperative communication and attribution processes to provide a comprehensive framework to categorize different types of incivility based on recipients' assessment of cooperative communication. An online experiment ( $N = 374$ ) was conducted in which participants were exposed to either one of five different types of incivility or a civil comment. It was explored how attributions about the sender's (1) intentions to discuss, (2) tolerable political views, and (3) discussion skills differ between the experimental conditions. Moreover, we explored how the distinct types of incivility affect individuals' behavior. Results demonstrated that attributions about senders differed according to the type of incivility participants were exposed to and, therefore, provide a novel opportunity to categorize incivility on a perceptual level. Foremost, disrespect towards others stood out on all attributional dimensions as uncooperative. However, the effect of the experimental conditions on participants' discussion behavior was limited.

*Keywords:* Incivility, online discussions, cooperative communication, attribution theory, user comments

Due to its negative impact on discussion dynamics, incivility is usually considered a serious challenge for online political discussions (e.g., Gervais, 2015; Ziegele, Jost, et al., 2018). However, incivility research faces some obstacles. Foremost, definitions and operationalizations of incivility are often inconsistent. Consequently, different studies on incivility are only comparable to a limited extent. Moreover, recent studies often highlight the subjective nature of incivility and conceptualize incivility as a perceptual construct that can manifest in different ways (e.g., Kenski et al., 2020; Muddiman, 2017; Stryker et al., 2016). Hence, Chen (2017) argues that incivility lies on a “continuum of non-normative speech that ranges from impoliteness to virulent hate speech” (p. 84). Although scholars have shifted to multidimensional and differentiated concepts of incivility, experiment-based studies usually consider only one type of incivility or mix different types of incivility when examining how incivility affects discussion dynamics (Chen & Lu, 2017; Hutchens et al., 2019; Rösner et al., 2016). As different types of incivility are conceptually distinct, it is important to examine whether they have different effects, which have not yet been scrutinized. Studies on perceptions of incivility suggest that distinct types of incivility are not equally harmful (e.g., Kenski et al., 2020; Stryker et al., 2016). Yet little is known about the underlying mechanisms that lead to different consequences of distinct types of incivility. Therefore, we introduce a novel approach to measure and categorize incivility based on recipients’ attributions about the sender and the resulting discussion behavior. Drawing on a more recent approach to incivility (Bormann et al., 2021), we conceptualize incivility as perceived violations of norms of cooperative communication. Hence, based on hypotheses about cooperative communication (Allwood et al., 2000; Tomasello et al., 2005) and assumptions of attribution theory (Malle, 1999), we explored how attributions about intentions to discuss, tolerable political views, and discussion skills vary among the distinct types of incivility. Additionally, we investigated how recipients’ discussion behavior is affected by the distinct types of incivility. This way, we aim (1) to highlight the role of social perception within the discussion about incivility, (2) to

provide new opportunities to measure incivility on a perceptual level, and (3) to offer new perspectives on how distinct types of incivility can be categorized based on recipients' assessments about cooperation.

### **Literature review**

Although incivility is not a new phenomenon, there is an ongoing debate on which behavior can be considered uncivil and whether incivility is inherently bad (Chen et al., 2019). A common denominator of different studies is that they usually conceptualize incivility as communicative behavior that violates norms (e.g., Chen, 2017). However, there is disagreement about which norms constitute uncivil communication. Here, the distinction between violations against interpersonal politeness norms and democratic norms is most prominent. While some scholars approach incivility as impoliteness (e.g., Chen & Lu, 2017; Mutz, 2007), other scholars such as Papacharissi (2004) postulate that violations of politeness norms are not per se detrimental to discussions but can even foster open exchange. In contrast, she conceptualizes incivility as a violation of democratic norms (i.e., verbalizing threats to democracy or individual rights and assignments of stereotypes) and argues that only such norm violations negatively affect the democratic discourse. Muddiman (2017), on the other hand, integrates both concepts and distinguishes between personal-level incivility, which corresponds to conceptualizations of impoliteness, and public-level incivility, which is conceptualized as violations against political and deliberative norms.

Thus, the most prominent dimensions of incivility in the extant literature are violations of politeness norms and democratic norms. As the politeness norm is usually violated on an interpersonal level, it can be considered a threat to people's faces, which can be defined as one's socially desired public image (Brown & Levinson, 1987; Chen & Lu, 2017). In contrast, the violation of democratic norms is rather immoral towards (often not present) others compared to face-threatening and is often associated with ideologic opinion expression (e.g., Rossini, 2019, 2020).

In addition, literature on incivility also highlights non-normative communication behaviors beyond the violation of these norms. For instance, Stryker and colleagues (2016) found that people perceive deceptive behavior as uncivil as well. Likewise, other discussion behaviors such as the usage of sarcasm or topic deviation have been linked to incivility within literature (e.g., Anderson & Huntington, 2017; Hopp, 2019; Ziegele & Jost, 2020). Although it has not yet been tested whether people perceive such behaviors as uncivil, categories of incivility seemingly go beyond the violation of interpersonal politeness norms and democratic norms.

Although recent studies have moved to a broader approach to incivility, a comprehensive theoretical framework that justifies why different types of non-normative communicative behavior can be subsumed under the umbrella of incivility is still missing. So far, most approaches addressing this issue have focused on differences in the manifestation of non-normative speech acts (e.g., Muddiman, 2017; Papacharissi, 2004; Rossini, 2020). However, when considering incivility as a multidimensional construct encompassing a range of violations against communicative norms (e.g., Chen, 2017), a theoretical fundament is needed that focuses on the common roots of the distinct behaviors rather than their differences. An approach, recently introduced by Bormann et al. (2021), takes this into account by conceptualizing incivility as a perceived violation of distinct communication norms that facilitate cooperative communication.

Thus, as incivility can be approached as a perceptual construct (see also Chen et al., 2019; Muddiman, 2017; Stryker et al., 2016), one suggestion is that incivility should be assessed by its consequences rather than by its mere manifestation (Chen et al., 2019). However, to assess incivility in this way, we need to find new ways to (1) measure incivility on a perceptual level, and we need to (2) generate more knowledge about the consequences for recipients of distinct types of incivility. Only then, meaningful categories of incivility can be derived, which allow predictions about their potential outcomes. The most (at least

immediate) negative outcomes are those which have a direct negative impact on the discussion dynamics. Therefore, we argue that a theoretical fundament of incivility can be derived from assumptions about cooperative communication.

### *1.1. Cooperative communication*

Although discussing politics online might not be the first activity that comes to mind when thinking of cooperative actions, such discussions require a high degree of mutual coordination and consideration. Discussions thrive on interaction and reciprocal exchange between people and, hence, are a highly social activity – even if people have different opinions. Allwood and colleagues (2000) argue that communication between two or more parties is basically cooperative as long as the parties cognitively consider each other and have the joint purpose of understanding each other (see also earlier conceptualizations by Grice, 1975). This corresponds to the hypothesis of *shared intentionality* as proposed by Tomasello and colleagues (2005). The authors argue that only human beings can form shared intentionality to act cooperatively. This ability does not only involve an understanding of *joint intentions* and *shared goals* but also a *mutual cognitive representation* of each other's skills and actions.

However, for *ideal cooperation*, Allwood and colleagues (2000) added two further requirements: *ethical consideration* and *trust*. Ethical consideration means that communication partners each have ethical goals, make ethical decisions, and act ethically. With trust, Allwood and colleagues (2000) do not refer to a holistic trust in others, but the trust that someone will act following the other requirements. Thus, ethical consideration and trust are strongly related to the basic requirements for cooperation: cognitive consideration and joint purpose (Allwood et al., 2000).

Therefore, cooperative communication in online discussions can be hindered in different ways. When people do not trust others to have an intention to discuss or to consider others cognitively, cooperative communication probably fails at a very basic level. Moreover,

when people do not trust others' ethical considerations, basic cooperation might develop, but the cooperation will probably not proceed in an ideal way.

### *1.2. Incivility as a signal for uncooperative communication*

A recent approach by Bormann and colleagues (2021) suggests that all types of incivility root in violations of norms that enable cooperative communication. Indeed, whether cooperative communication succeeds strongly depends on individuals' trust in others to act following the requirements proposed by Allwood and colleagues (2000). Here, an obvious assumption is that incivility can decrease this trust and consequently impedes cooperative communication. Indeed, scholars identified various negative outcomes of incivility that support this assumption. For instance, Han and Brazeal (2015) found that people are less willing to participate in online discussions when exposed to incivility. Likewise, Han and colleagues (2018) found that exposure to incivility decreases the likelihood that people respectfully acknowledge others' perspectives, people stay on the discussion topic, and add new perspectives to the discussion. Although individuals who were exposed to incivility have not been more likely to communicate uncivilly themselves in this study, other studies found that incivility triggers additional uncivil commenting (Chen & Lu, 2017; Gervais, 2015, 2017; Ziegele, Weber, et al., 2018). In sum, these results indicate that incivility and uncooperative discussion behavior are related in important ways. Therefore, the lens of cooperation helps to understand why and when incivility has detrimental effects on discussion dynamics. Based on the framework provided by Allwood and colleagues (2000), it can be suggested that such behavior results from a lack of trust in the sender's willingness and/or ability to cooperate. Thus, incivility might be better categorized by perceptions about cooperative communication instead of the mere usage of a certain wording.

In this vein, following the assumptions of Bormann and colleagues (2021), we consider that a linguistic perspective might also prove to be helpful in explaining why people appraise certain antinormative behavior to be uncivil that goes beyond the violation of

politeness norms or democratic norms: Grice's (1975) cooperation principle suggests that people expect that conversation partners make their contribution "such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which [they] are engaged" (p. 45). He introduced four categories of maxims that ensure that the principle of cooperation is fulfilled. The categories have been labeled *quantity* (i.e., "Make your contribution as informative as is required" but "not more informative than it is required"), *quality* (i.e., "Try to make your contribution one that is true"), *relation* (i.e., "be relevant"), and *manner* (i.e., "be perspicuous" p. 45 – 46). Although these maxims seem to be prescriptive, Grice (1975) rather describes what people expect from others when they communicate. In other words, Grice (1975) elucidates communicative norms. If people strongly violate these norms, it probably leads to a perceived lack of willingness and ability to cooperate and to negative implications about the relationship between the violator and the recipient (Mooney, 2004). Then, just like for violations against politeness norms or democratic norms, the (ideal) cooperation as outlined by Allwood and colleagues (2000) is hindered.

Thus, from Grice's (1975) cooperation principle and its maxims, we can derive categories of incivility that are less highlighted in the literature on incivility. First, following the maxims of quality and quantity (i.e., be as informative as required and tell the truth), we suggest that people consider severe norm violations that focus on information sharing as uncivil since inaccuracy can hamper cooperative communication. Here, communicating misleading information might be the most prominent norm violation, also reflecting the deception category identified by Stryker et al. (2016). Moreover, following the maxim of relation (i.e., be relevant), in online discussions, people are supposed to stick to the discussion topic and communicate reciprocally; otherwise, no mutual exchange, and thus no discussion can arise. Therefore, it is not surprising that some studies have conceptualized topic deviations, interruptions, and ignoring other participants' contributions as types of incivility

(e.g., Hopp, 2019; Stryker et al., 2016; Sydnor, 2018). Although Grice (1975) noted that there are cases in which it can be legitimate to change the topic of conversation, in (online) political discussions, an unjustified departure from the topic of interest is probably perceived as uncooperative behavior. Regarding the maxims of manner (i.e., be perspicuous), irony and sarcasm are common stylistic devices in discussions. Although not necessarily intended to be ambiguous, the usage of irony can lead to misunderstandings, and sarcasm is often additionally mocking. Therefore, we consider ironic sarcasm to be a representative type of incivility when violating norms of modality. Even though this type of incivility might also work on an interpersonal politeness norm, we think that the fact that someone intentionally expresses oneself ambiguously is the more salient norm violation.

Taken together, communicative contributions can be considered a vehicle to signal cooperation. Conversely, this means that a lack of willingness or ability to cooperate can already be signaled at a very basic linguistic level by violating inherent communication norms.. Thus, incivility can be defined as *perceived violations of norms of cooperative communication* (Bormann et al., 2021). This way, different types of incivility can be categorized by the way they negatively affect cooperative communication on a perceptual level. As can be seen in Table 1, different types of incivility can manifest in the violations of distinct communication norms.

Table 1

*Overview of distinct communicative norms and distinct types of incivility as norm violations.*

Communicative Norms	Representative norm violation
Information norm	Spreading misleading information (e.g., Stryker et al., 2016)
Norms of manner	Using irony or being sarcastic (e.g., Anderson & Huntington, 2017)
Norms of relation	Deviating from the discussion topic (e.g., Hopp, 2019)
Politeness norms	Disrespecting other participants (e.g., Coe et al., 2014)
Democratic norms	Violations of democratic norms such as negative stereotyping of groups or threats of violence (e.g., Papacharissi, 2004)

However, it is important to understand the ways people interpret others' communication behavior and which inferences they make when observing it. Here, online environments are especially challenging. Participants only have limited cues to understand what the reasons for others' behavior are. When individuals want to discuss with others online, they usually have to decide whether someone is an appropriate discussion partner solely based on the appearance of their comments.

### *1.3. Attributions in cooperative discussion behavior*

The process of how people interpret others' behavior is addressed by attribution theories. These theories aim to answer the question of how people make sense of other people's behavior. When people explain others' behavior, they sharply distinguish between intentional (e.g., lie down) and non-intentional behavior (e.g., fall down; Malle, 2011). When asking whether a comment is written on purpose, probably everyone would agree that people do not write user comments by accident but that they write comments quite intentionally. However, this question becomes more complex when asking why a person wrote a comment in a certain form (e.g., in an uncivil way).

Following Malle's (1999) *Folk-Conceptual Theory of Behavior Explanation*, people rely on distinct modes of explanation when they assess others' behavior as intentional. The most important modes are *reason explanations* and the *causal history of reason explanations*. In a nutshell, reason explanations represent the intentions that are inferred from actors. The mode of the causal history of reason explanations addresses factors that lie beyond individuals' intentions. This comprises, for example, the observed actor's personality, the person's cultural background, as well as the person's unconscious mental state. The third mode, *enabling factors*, is described as a rare mode of explanation as it does not provide explanatory power for the actor's motivation but only clarifies how it was possible that the intended behavior was (not) performed.

The Folk-Conceptual Theory of Behavior Explanation by Malle (1999) has much in common with theories and analyses addressing cooperation (Allwood et al., 2000; Grice, 1975; Tomasello et al., 2005) since people have to derive attributions about other's intentions and personal characteristics from the cooperation partner's behavior. Despite this strong overlap, these two lines of research have, so far, barely been brought together. However, a combination of these frameworks might be fruitful to explain how people make sense of others' (uncivil) discussion behavior. When people perceive certain behavior as uncooperative, individuals can derive the reasons from different dimensions: *Are certain personal characteristics central for uncooperative behavior? Does the person not want to cooperate? Or is the person not able to cooperate?*

Thus far, literature fundamentally lacks knowledge on how different types of incivility affect attribution processes and, more specifically, attributions about the potential cooperativeness of others. Only when individuals assume that an uncivil commenter has the purpose of discussing will they probably expect further discussion potential (Allwood et al., 2000). Nevertheless, the other modes of explanation as proposed by Malle (1999) can be important, too. In online political discussions, various factors can affect the attribution

process. Especially the political stance and attitudes toward certain topics seem to be important in online discussions. Even when people think a person has the desire to discuss, the mode of the causal history of reason explanation might indicate that the underlying factors are not ethical and, consequently, the person will not consider the other person as an appropriate discussion partner (Allwood et al., 2000). Although this attribution is also likely to be related to the attribution of the persons' intentions, this inference might be more important for assessing the cooperative potential. Moreover, people must be convinced that people have a certain ability to discuss a specific topic with others. That is important because people can be convinced that others have ethical attitudes and ethical intentions, but at the same time, they can doubt the discussion skills of these persons and, therefore, are less likely to see them as appropriate discussion partners. Of course, people probably do not always make these attributions consciously. However, to some degree, people must be convinced that others cognitively and ethically consider them in the discussion and have the intention to discuss to recognize them as discussion partners (see Allwood et al., 2000; Tomasello et al., 2005).

Overall, an attribution theory approach clarifies how people process uncivil behavior in terms of ascribing cooperation potential. However, as this perspective has been neglected in research, little is known about the attributions people make when others behave uncivilly in different ways.

#### *1.4. Different types of incivility – different types of attributions?*

Although it is rather subjective which discussion behaviors are uncivil and which are acceptable (e.g., Chen et al., 2019), certain standards can be applied to public discourse. In public online discussions (e.g., in discussion forums or comment sections on social media), there are usually guidelines (e.g., netiquettes) indicating that violations of politeness and democratic norms are not acceptable, or asking participants to stay on topic and to discuss reciprocally. Moreover, as people inherently expect that people communicate in accordance

with Grice's (1975) cooperation principle, strong violations of his maxims will probably be perceived as such. Therefore, in the context of common public discussions, we expect that people who violate norms are perceived as less cooperative. More precisely, following Allwood et al. (2000), if people regard others' attitudes or commenting intentions as unethical, they are probably less likely to see them as discussion partners. Likewise, people must have a certain trust in the commenter's skills to discuss.

Against this background, we argue that violations of communicative norms of cooperation affect the attribution process as they can be interpreted as signals for violators' uncooperative characters; and thus are considered uncivil. Therefore, if people are exposed to comments that include such signals, we expect that they attribute fewer discussion intentions to the sender of the message, less tolerable political views, and fewer discussion skills than to senders of comments without such signals. We expect that this applies to norm violations as derived by Grice (1975): information norms, norms of manner, and norms of relation. Likewise, we expect that this applies to interpersonal politeness norms, and democratic norms (e.g., Muddiman, 2017; Papacharissi, 2004):

*H1: Compared to civil commenting behavior, violations against (1.1) information norms, (1.2) norms of manner, (1.3) norms of relation, (1.4) interpersonal politeness norms, and (1.5) democratic norms negatively affect recipients' attributions of the violator's (a) intentions to discuss, (b) tolerable political views, and (c) discussion skills.*

Furthermore, the question arises of how the assessment of the distinct types of norm violations differs regarding these attributions. So far, little work has been done on this. Since all of the proposed norm violations have unique features, we expect that they are processed differently. For violations of politeness norms, it was already found that using vulgar language and name-calling (compared to norm-compliant comments) leads to stronger attributions of aggressive discussion motives and decreased attributions of interest in others' opinions, indicating that people attribute less discussion intention to commenters who violate

against incivility in the form of politeness norms (Kluck & Krämer, 2021). Therefore, one could argue that violations against interpersonal politeness norms decrease attributions of cooperative intentions in the form of discussion intentions. However, Papacharissi (2004) argues that impoliteness can even foster the exchange in discussions. Accordingly, Rossini (2020) found that such behavior is often related to justified opinion expressions. Therefore, it is difficult to predict to which degree people attribute cooperative intentions (i.e., discussion intentions) to people who violate interpersonal politeness norms. Moreover, since violations of politeness norms are often associated with emotionality (e.g., Chen & Ng, 2017; Gervais, 2015, 2017), people might attribute limited discussions skills (see also Kluck & Krämer, 2021) to the violator as they attribute a low amount of self-control to them. Likewise, which attributions are related to violations of democratic norms is difficult to determine. Such violations are often characterized as immoral acts (e.g., Rossini, 2020). However, Rossini (2020) found that such norm violations in the form of politically intolerant behavior are most likely to spread in ideologically homogeneous spaces. Therefore, if people politely violate democratic norms, people might attribute intentions to discuss to the violator.

Also, little is known about the attributions people make when commenters violate more basic communicative norms of cooperation like those derived from Grice's (1975) maxims. Since individuals inherently assume that others communicate according to the principle of cooperation, they usually do not consider violations of the related communicative norms to be serious. Rather they are searching for the implicature of the message (Grice, 1975). Nevertheless, if these violations are more severe, they will probably also have an impact on the perceived cooperation. However, again, the attributions which are most likely to be affected by violations of basic communication norms are difficult to predict. As people usually search for the meaning of a message rather than focusing on linguistic norm violations (Grice, 1975), these norm violations might strongly influence which intentions are attributed

to the violator. On the other hand, people might lose trust in the violator's discussion skills, as basic conversational norms are violated that are usually expected to be followed.

In sum, it is unlikely that the distinct norm violations lead to the same attributions about the violator. However, as this research issue has been scarcely addressed so far, no hypotheses can be derived about how the different types differ concerning the attributions of the recipients. Therefore, we expect and ask:

*H2: Attributions about the violator's (a) intentions to discuss, (b) tolerable political views, and (c) discussion skills will differ among the distinct norm violations.*

*RQ1: In which ways do these attributions differ among the distinct norm violations?*

### *1.5. Behavioral outcomes*

Similar to the attribution process, little is known about how the different types of norm violations lead to different cooperative discussion outcomes. Some types of incivility have already been found to have negative effects on the discussion behavior of its recipients. For instance, Han and Brazeal (2015) found that incivility leads to less agreement, discourages individuals from adding new perspectives to the discussion, and discourages them from discussing at all. Likewise, Han and colleagues (2018) found that people are less likely to add a new perspective to the discussion when being exposed to uncivil comments and are less likely to stay on the discussion topic. However, these studies did not distinguish between different types of incivility but mixed distinct types of norm violations.

As studies that examined behavioral outcomes rarely differentiate between distinct types of incivility, we expect that the derived norm violations have different effects on individuals' behavior. Since the most detrimental effect of incivility probably is that people do not discuss at all, it is important to examine which type of incivility has the most potential to have such an effect. Further, to be constructive, participants of a discussion need to stay on the discussion topic, and contributions need to be argumentative (see also Han et al., 2018).

All of these aspects are particularly important when participants answer the uncivil comment.

Therefore, we further ask:

*RQ2a-d: Which norm violations have the most detrimental effect on the receiver's likelihood (a) to stay in the discussion, (b) to answer to the norm violating comment, (c) to stay on the discussion topic when answering the norm violating comment, and (d) to add arguments to the discussion when answering the norm violating comment?*

Moreover, whereas some researchers found that incivility leads to more incivility (Chen & Lu, 2017; Gervais, 2015, 2017; Ziegele, Weber, et al., 2018), other researchers could not support this finding (e.g., Han & Brazeal, 2015; Rösner & Krämer, 2016). The conflicting findings can partly be explained by the fact that most of these studies coded uncivil reactions differently and did not systematically distinguish between different types of incivility. Thus, when differentiating distinct types of incivility, specific patterns might emerge:

*RQ3: Do the distinct norm violations trigger further uncivil behavior, and are there differences between distinct types of incivility?*

Based on our theoretical argumentation, it can be expected that the attributions about the norm violator's cooperativeness affect the behavioral outcomes. However, as no differential effects can be derived, we finally ask:

*RQ4: Which role do attributions regarding the violator's cooperativeness play in different behavioral outcomes?*

## **2. Method**

To investigate the hypotheses and research questions, an online experiment was conducted from 6 to 22 April 2021, addressing a German-speaking population. Participants were requested to participate in an ostensible real online discussion. Within the discussion, individuals were either exposed to a civil discussion comment or a comment including one type of incivility (i.e., stating false information, deviating from the discussion topic, using

ironic sarcasm, being disrespectful to other discussants, or violating democratic norms). Analyses were conducted with IBM SPSS 26. The study was preregistered<sup>1</sup> and approved by the local [blinded for Review] Ethics Review Board. Informed consent was obtained from all individual adult participants included in the study; children were not allowed to participate.

### 2.1. *Sample*

In total, 393 participants were recruited from a commercial online panel ( $n = 272$ ) and a non-commercial platform for scientific surveys ( $n = 121$ ). Sixteen cases have been excluded from analyses based on suspicious response behavior, identified through attention checks and a manual data check. The remaining 374 individuals (189 females, one diverse, and one not specified) were aged between 18 and 68 ( $M = 38.01$ ,  $SD = 14.33$ ). Most participants indicated being employed (47%) or being university students (29%). Moreover, most participants reported having a university degree (39%). Around 30% of participants indicated that they take part in social media discussions daily or several times a week, whereas around 51% said they discuss less frequently than once a month on these platforms. However, the majority stated that they read social media discussions daily or several times a week (60%).

### 2.2. *Design and Procedure*

An ostensible discussion platform with the name "Let's discuss" was created using a no-code website builder. Within a separate survey, participants were exposed to the study's cover story. They were told that the study is about testing a new kind of discussion forum. It was claimed that the forum has innovative rules allowing a more balanced discussion. In fact, these rules were developed to control the sequence of the discussion posts and to ensure that the participants had to react to a specific comment that was manipulated according to the experimental conditions. In this way, a situation was created in which communicative cooperation was requested. The rules claimed that the discussion is turn-based and that discussants should always address the previous comment to ensure that no opinion is

---

<sup>1</sup> [https://osf.io/p8z6u/?view\\_only=ad81ffd6d58742cf919b69e1ac128f9c](https://osf.io/p8z6u/?view_only=ad81ffd6d58742cf919b69e1ac128f9c)

neglected. It was also explained that if discussants cannot or do not want to reply to the previous comment, they can pass one round.

When participants were directed to the external discussion forum, they were equally allocated to one out of two fixed topics (i.e., the expansion of referendums in Germany or the introduction of an unconditional basic income in Germany). These two topics were merged in the analyses to increase the generalizability of the results. Within the discussion, some information about the discussion topic has been provided to ensure that all participants have the same basic knowledge about the issue. Comments were displayed one after the other, and it took on average 39 seconds until a comment was posted. For all participants, the order of comments was the same. The first comment indicated an ambivalent attitude towards the discussion topic; the second comment indicated a con position towards the discussion topic; the third comment indicated a pro position towards the discussion topic. Participants always answered last and were supposed to react to the third comment according to the forum's rules. The third comment was manipulated regarding the experimental condition and was either civil or included one type of incivility: Stating false information, deviating from the discussion topic, using ironic sarcasm, being disrespectful to other discussants, or stereotyping and threatening social groups.

When the last comment was displayed, participants had the chance to react. They had the opportunity to write an answer, pass a round, or leave the discussion. If they chose to pass a round or leave the discussion, they were redirected to the survey – ostensibly to answer some questions about their experience of participating in the first round of discussions. If participants chose to write an answer, they could do so and were redirected to the survey afterward. First and in a random order, participants answered questions about the discussion intentions, the general attitudes, and the discussion skills of the preceding commenter. Last, they evaluated the commenter's warmth and competence as well as the extent of aggressive discussion motives. Likewise, participants were asked about their perception of the comment.

### 2.3. *Experimental manipulation*

To ensure that the different types of incivility are perceived as such, we pretested all comments. Please see online Appendix A for the results of the pretest. Except for the condition of topic deviation, for each topic, all manipulated comments included the same core argument and addressed the previous commenter. However, the comment's form was manipulated according to the different types of incivility. Within the condition of false information, the comment included additional untrue claims and misleading exaggerations. The ironic comment apparently stated the opposite of the sender's actual opinion, which was marked by an overemphasis of counterintuitive arguments typical for irony. The comment deviating from the discussion topic neither included any argument regarding the subject of discussion nor – against the forum's rules – addressed the previous commenter. Instead, it was implied that another topic is more important. The comment being disrespectful to other discussants was offensive to the previous commenter and included profane language. Finally, the comment with violations of democratic norms included antagonistic stereotyping of and a call for violence against certain social groups (in both cases, elites). Please note that a comprehensive description of the forum, its features, and the manipulation of the comments can be found in the online Appendix B.

### 2.4. *Measures*

**3.4.1. Attributions.** Since no standardized measurements exist which would capture the attribution variables, we constructed our own scales with five items each (7-point Likert scales from 1 = *do not agree at all* to 7 = *totally agree*). To measure which extent of (1) *discussion intentions* participants attribute to the (un)civil sender, participants rated statements such as “The person wants to discuss with the other participants” or “The person wants to bring new perspectives to the discussion” (Cronbach's  $\alpha = .92$ ). To measure the attribution of (2) *tolerable political views*, participants were informed that they should assess whether the preceding commenter's political views are basically tolerable and whether they agree with

these views. Exemplary items are “Most people would tolerate the person's political views” or “The person has political views that are basically acceptable” (Cronbach’s  $\alpha = .91$ ). The attribution of (3) *discussion skills* was measured with items like “The person can formulate contributions in such a way that she/he achieves the intended goal” or “The person can formulate comments in a way that does not create misunderstandings” (Cronbach’s  $\alpha = .94$ ). For validation of the measurements, please see Appendix C.

**3.4.2. Discussion behavior.** To investigate individuals’ reactions, we analyzed their actual discussion behavior. Participants’ willingness to stay in the discussion and to answer the preceding comment was assessed by the usage of the buttons “pass round” and “leave the discussion.” If the corresponding button was used, it was coded as 1 and otherwise as 0. If participants used one of the buttons, they were not able to write a comment anymore. To examine the commenting behavior of those who decided to respond ( $n = 314$ ,  $M = 37.89$  words  $SD = 19.39$ , word range: 5 – 111), answers were coded by two authors of the study. A codebook was written to define the different categories and the authors trained the coding scheme based on 15 comments. Afterward, 60 comments were coded by both researchers to examine the intercoder reliability measured by Krippendorff’s alpha. The remaining comments were each coded half by one author.

It was examined whether the comments addressed the discussion topic (i.e., explicitly addressing the discussion topic or an aspect of the topic;  $\alpha = .84$ ) and whether comments contained an argument (i.e., giving justifications for opinions and statements;  $\alpha = .86$ ). Additionally, although not indicated in the preregistration, we coded whether the comments criticized the preceding comment. Following Gervais (2015), critique is a way to signal aversion toward other discussants by negatively assessing “specific personal qualities, behavior, and traits” of them ( $\alpha = .73$ ).

Moreover, it was investigated which types of incivility have been used in the comments. Therefore, based on the employed manifestations of incivility in the stimulus

material, it was coded whether the comment included (1) misleading information (i.e., dubious factual allegations or misleading exaggerations;  $\alpha = .70$ ), (2) irony or sarcasm (i.e., clear ironic or sarcastic statements;  $\alpha = .73$ ), (3) derailing behavior (i.e., in no way connected to the previous discussion;  $\alpha = .71$ ), (4) violations against respect or politeness norms towards other discussants (e.g., insults, name-calling, vulgarity, profanity, degradation;  $\alpha = 1.00$ ), (5) violations of democratic norms (e.g., negative stereotyping of or threats against certain groups, threats to democracy;  $\alpha = .79$ ).

**3.4.3. Manipulation check.** To check whether participants recognized the distinct norm violations as intended, for each norm violation, participants rated on one item (7-point Likert scale from 1 = *do not agree at all* to 7 = *totally agree*) whether the comment included the respective features (e.g., “has personally attacked a person present in the discussion” for disrespect towards others).

### 3. Results

#### 3.1. Manipulation check

To check whether individuals recognized the respective norm violations, we conducted a series of t-tests. This involved comparing the mean values of the one-item measures assessing the presence of each norm violation between the civil condition as a baseline and the condition with the relevant norm violation. As provided in Table 2, all norm violations have been assessed as such by participants providing large effect sizes. Noteworthy, misleading information seemed to be the most difficult for participants to identify.

Table 2.

*T-tests checking whether manipulations of the comments were perceived in the intended direction.*

Item asking for...	Civil condition		Condition with respective norm violation		<i>t</i> (df)	<i>p</i>	<i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
misleading information	1.84	1.55	3.77	2.13	- 5.85 (114,97)	< .001	1.04
ironic sarcasm	1.36	0.98	4.09	2.55	-8.01 (81,29)	< .001	1.41
topic deviation	1.55	1.08	4.28	1.83	-10.04 (94,46)	< .001	1.81
disrespect towards others	1.25	0.89	4.71	2.59	-9.74 (70,54)	< .001	1.82
violation of democratic norms	1.47	1.13	4.06	2.38	-7.83 (88,10)	< .001	1.39

### 3.2. *Attributions*

We expected that compared to civil comments, all types of incivility decrease attributions of recipients' attributions of the violator's (H1.1a – H1.5a) intentions to discuss, (H1.1b – H1.5b) tolerable political views, and (H1.1c – H1.5c) discussion skills. Moreover, we expected (H2) that there are differences between the distinct types of incivility regarding the dependent variables and asked (RQ1) what patterns emerge.

As Levene's test indicated that variances were unequal across the groups for all dependent variables, we conducted ANOVA's employing Welch's *F* and Bonferroni post hoc tests to examine differences between the groups (Field, 2013). See *Table 3* for all means and standard deviations of the dependent variables.

Table 3

*Means and standard deviations of the different attributions for each condition.*

Conditions	Attributions					
	Discussion intentions		Tolerable political views		Discussion skills	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
(a) Civil ( <i>n</i> = 64)	27.83 <sup>c,d,e,f</sup>	5.61	28.31 <sup>c,e,f</sup>	5.35	27.97 <sup>c,e</sup>	5.85
(b) Misleading information ( <i>n</i> = 64)	25.98 <sup>d,e</sup>	5.71	25.86 <sup>e,f</sup>	5.54	27.00 <sup>c,e</sup>	5.45
(c) Ironic sarcasm ( <i>n</i> = 64)	22.30 <sup>a,e</sup>	7.16	24.07 <sup>a,f</sup>	6.39	21.61 <sup>a,b</sup>	8.57
(d) Topic deviation ( <i>n</i> = 60)	21.73 <sup>a,b,e</sup>	7.28	25.93 <sup>e,f</sup>	5.28	25.15 <sup>e</sup>	6.70
(e) Disrespect to others ( <i>n</i> = 59)	17.51 <sup>a,b,c,d,f</sup>	8.68	21.07 <sup>a,b,d</sup>	7.15	20.00 <sup>a,b,d,f</sup>	8.84
(f) Violation of democratic norms ( <i>n</i> = 63)	23.10 <sup>a,e</sup>	9.01	20.11 <sup>a,b,c,d</sup>	8.87	25.11 <sup>e</sup>	7.05

*Note.* Superscript letters indicate that the mean values are significantly different from the conditions prefixed with the same letter ( $p < .05$  based on Bonferroni post hoc tests).

**4.2.1. Discussion intentions.** Regarding discussion intentions, results revealed that there was a significant difference between the experimental groups (Welch's  $F(5,169.97) = 15.85$ ,  $p < .001$ ,  $\omega^2 = .16$ ). Bonferroni post hoc tests indicated that the civil condition significantly differed from conditions including ironic sarcasm ( $p < .001$ ), topic deviation ( $p < .001$ ), disrespect to other discussants ( $p < .001$ ), and violations of democratic norms ( $p = .005$ ). When exposed to misleading information, attributions of discussion intent significantly differed from topic deviation ( $p = .020$ ) and disrespect to other discussants ( $p < .001$ ). Moreover, attributions of discussion intentions were significantly for disrespect towards other discussants than for ironic sarcasm ( $p = .005$ ), topic deviation ( $p = .027$ ), and violations of democratic norms ( $p < .001$ ). However, no significant differences were found between the civil condition and misleading information ( $p > .999$ ); ironic sarcasm and topic deviation ( $p > .999$ ) as well as misleading information ( $p = .070$ ). Likewise, no differences between democratic norms and misleading information ( $p = .405$ ), ironic sarcasm topic ( $p > .999$ ), and topic deviation emerged ( $p > .999$ ). Thus, in sum, the assumptions that the distinct types of

incivility lead to decreased attributions of discussion intentions have largely been supported (H1.2a – H1.5a). However, contradicting H1.1a, the exposure to misleading information did not lead to decreased attributions of discussion intentions. Supporting H2a, the effects on the attribution of discussion intentions differed among the distinct types of incivility. Foremost, disrespect to other discussants decreased attributions of discussion intentions (*RQ1*).

**4.2.2. Tolerable political views.** For attributions of tolerable political views, again, significant differences have been revealed (Welch's  $F(5,170.65) = 13.03, p < .001, \omega^2 = .15$ ). Compared to the civil condition, individuals attributed less tolerable political views to a commenter who applied ironic sarcasm ( $p = .004$ ), disrespect to other discussants ( $p < .001$ ), or violations against democratic norms ( $p < .001$ ). Additionally, significant differences emerged between disrespect and misleading information ( $p = .001$ ), topic deviation ( $p = .001$ ), and violations against democratic norms ( $p < .001$ ). Likewise, violations of democratic norms differed significantly from misleading information ( $p < .001$ ), ironic sarcasm ( $p = .011$ ), and topic deviation ( $p < .001$ ). In contrast, no differences emerged between the civil condition and the usage of misleading information ( $p = .522$ ) and topic deviation ( $p = .660$ ). Likewise, misleading information did not differ from ironic sarcasm ( $p > .999$ ) and topic deviation ( $p > .999$ ) regarding attributions of tolerable political views. Also, disrespect to other discussants did not differ significantly from ironic sarcasm ( $p = .170$ ) and violations of democratic norms ( $p > .999$ ).

As the exposure to ironic sarcasm decreased attributions of tolerable political views, and as individuals attributed significantly less tolerable political views to commenters showing disrespect to others or violating democratic norms, H1.2b, H1.4b, and H1.5b were supported. However, as the exposure to misleading information or topic deviation did not influence the attributions about the commenter's tolerable political views, H1.1b, and H1.3b had to be rejected. Moreover, regarding H2b and RQ1, especially disrespect to others and violations of democratic norms diminished attributions of tolerable political views.

**4.2.3. Discussion skills.** Examining the attribution of discussion skills, significant differences have been found, too (Welch's  $F(5,170.18) = 10.35, p < .001, \omega^2 = .12$ ). The civil condition differed significantly from the conditions with ironic sarcasm ( $p < .001$ ) and disrespect towards others ( $p < .001$ ). Likewise, when exposed to misleading information, people attributed significantly more discussion skills than when exposed to ironic sarcasm ( $p < .001$ ) or disrespect towards other discussants ( $p < .001$ ). Additionally, when exposed to disrespect towards other discussants, participants attributed significantly fewer discussion skills than when exposed to topic deviation ( $p = .002$ ) or the violation of democratic norms ( $p < .001$ ). However, the civil condition did not differ significantly from the condition with misleading information ( $p > .999$ ), topic deviation ( $p = .440$ ), and violations against democratic norms ( $p = .380$ ). Moreover, misleading information and topic deviation as well as violations of democratic norms (both  $p > .999$ ) did not differ significantly whereas ironic sarcasm did not differ from topic deviation ( $p = .094$ ), disrespect to other discussants ( $p > .999$ ), and the violation of democratic norms ( $p = .093$ ). Likewise, topic deviation did not significantly differ from the violation of democratic norms ( $p > .999$ ).

Taken together, H1.1c, H1.3c, and H1.5c were not supported because neither misleading content, topic deviation, nor the violation of democratic norms decreased attributions of discussion skills compared to civil comments. However, H1.2c was supported as ironic sarcasm decreased attributions of discussion skills. Likewise, as comments including disrespect to others decreased these attributions compared to civil comments, H1.4c was confirmed. Again, supporting H2c and addressing RQ1, differences between the distinct types of incivility were found. Foremost, disrespect towards other discussants but also ironic sarcasm decreased the attribution of discussion skills.

### 3.3. Behavioral outcomes

To examine behavioral outcomes, we initially planned to calculate a series of logistic regression analyses. However, only a few cases of uncivil or non-constructive discussion

behavior emerged among the experimental conditions. Therefore, for the effects of the experimental conditions on participants' behavior, we focused on descriptive analyses. The Fisher-Freeman-Halton (FFH) Exact Test was employed to examine the resulting differences.

First, it was examined whether there were differences between the conditions concerning participants' tendency to write a comment (RQ2a&b). However, as presented in *Table 4*, FFH Exact Test revealed no significant differences in the number of individuals who used the "pass" ( $p = .795$ ) or the "leave" button ( $p = .795$ ). Additionally, we tested whether there are differences between the conditions when merging the two variables (i.e., generally refrain from discussing). However, again no differences were found ( $p = .795$ ).

Table 4

*The tendency of participants not to write a comment.*

$N = 374$	Civil ( $n = 64$ )	Misleading information ( $n = 64$ )	Ironic sarcasm ( $n = 64$ )	Topic deviation ( $n = 60$ )	Disrespect towards other discussants ( $n = 59$ )	Violation of democratic norms ( $n = 63$ )
Pass	4 (6.3%)	6 (9.4%)	8 (12.5%)	5 (8.3%)	7 (11.9%)	8 (12.7%)
Leave	1 (1.6%)	2 (3.1%)	5 (7.8%)	2 (3.3%)	6 (10.2%)	5 (7.9%)
No comment (Pass or Leave)	5 (7.8%)	8 (12.5%)	13 (20.3%)	7 (11.7%)	13 (22.0%)	13 (20.6%)

Next, differences between the conditions regarding participants' discussion behavior were explored (RQ2c-d). FFH Exact Test indicated that there were significant differences in terms of addressing the discussion topic ( $p < .001$ ). As presented in *Table 5*, the given discussion topic was addressed by 98.3% of participants in the civil condition, 98.2% in the condition of misleading information, 100% of participants in the condition of ironic sarcasm, 80.8% of participants in the condition of topic deviation, 89.1% in the condition of disrespect towards other discussants, and by 94.0% of participants who were exposed to violations of democratic norms. Regarding participants' use of arguments, FFH Exact Test indicated no

significant differences among the experimental conditions. However, differences emerged regarding expressions of critique (FFH Exact Test,  $p < .001$ ). Whereas in the civil condition, no cases of critique occurred, the preceding comment was criticized by 12.5% of participants in the condition of misleading information, 5.9% of participants in the condition of ironic sarcasm, 3.9% of participants in the condition including topic deviation, 34.8% in the condition including disrespect towards other discussants, and by 20.0% of participants in the condition with violations of democratic norms.

Table 5

*Participants' discussion behavior.*

$N = 314$	Civil ( $n = 59$ )	Misleading information ( $n = 56$ )	Ironic sarcasm ( $n = 51$ )	Topic deviation ( $n = 52$ )	Disrespect towards other discussants ( $n = 46$ )	Violation of democratic norms ( $n = 50$ )
Discussion Topic	58 (98.3%)	55 (98.2%)	51 (100.0%)	42 (80.8%)	41 (89.1%)	47 (94.0%)
Argument	51 (86.4%)	44 (78.6%)	47 (92.2%)	41 (78.9%)	38 (82.6%)	43 (86.0%)
Critique	0 (0.0%)	7 (12.5%)	3 (5.9%)	2 (3.9%)	16 (34.8%)	10 (20.0%)

Last, we investigated whether participants used different types of incivility based on the type of incivility they have been exposed to (RQ3). However, as presented in *Table 6*, only a few instances of uncivil behavior occurred at all. Overall, 23 cases of misleading information, 3 cases of sarcastic irony, 13 cases of topic deviation, 2 cases of disrespecting behavior, and 12 cases of violations against democratic norms were identified.

Regarding participants' usage of misleading information, no significant differences were found between the conditions (FFH Exact Test,  $p = .917$ ). Likewise, as only three instances of irony or sarcasm were identified, no significant differences were found (FFH Exact Test,  $p = .116$ ). However, FFH Exact Test indicated that there are differences between the conditions regarding individuals' tendency to deviate from the discussion content (i.e.,

addressing a new topic that was not included in the discussion before). 15.4% of participants who were exposed to a topic deviation, further derailed from the discussion by introducing a new issue that neither addressed the discussion topic nor the claim of the preceding comment. This was also true for 3.4% of participants in the civil condition and 6.0% in the condition including violations of democratic norms. In all other conditions, no cases of derailing emerged.

Regarding participants' disrespect towards other discussants, only two cases emerged in the condition in which they were exposed to the same behavior (3.9%). FFH Exact Test indicated that the difference was significant ( $p = .021$ ). However, based on the small number of cases, explanatory power is very limited. Regarding participants' violation of democratic norms, FFH Exact Test again indicated a significant difference among the different conditions ( $p = .029$ ). Democratic norms were violated by 5.1% of participants in the civil condition, 1.9% of participants in the condition of topic deviation, 5.8% of participants who were exposed to disrespect towards another discussant, and 10.0% of individuals who were exposed to earlier violations of democratic norms. No cases emerged in the conditions of misleading information or ironic sarcasm.

Table 6

*Participants' uncivil commenting behavior.*

<i>N</i> = 314	Civil ( <i>n</i> = 59)	Misleading information ( <i>n</i> = 56)	Ironic sarcasm ( <i>n</i> = 51)	Topic deviation ( <i>n</i> = 52)	Disrespect towards other discussants ( <i>n</i> = 46)	Violation of democratic norms ( <i>n</i> = 50)
Misleading information	5 (8.5%)	4 (7.1%)	2 (3.9%)	5 (9.6%)	3 (5.8%)	4 (8%)
Ironic sarcasm	1 (1.7%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	2 (3.9%)	0 (0.0%)
(Further) Derailing	2 (3.4%)	0 (0.0%)	0 (0.0%)	8 (15.4%)	0 (0.0%)	3 (6.0%)
Disrespect towards other discussants	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	2 (3.9%)	0 (0.0%)
Violation of democratic norms	3 (5.1%)	0 (0.0%)	0 (0.0%)	1 (1.9%)	3 (5.8%)	5 (10.0%)

*Note.* Rows represent participants' behavior.

For the behavioral variables for which differences were found between the conditions, we additionally investigated which role the different attributions regarding preceding commenters' discussion intention, tolerable political views, and discussion skills played (*RQ4*). Since the cases were generally low in all experimental groups, logistic regressions were performed separately with the different attributions as predictors.

Regarding participants' discussion behavior beyond incivility, it was found that most participants did not address the discussion topic when exposed to the condition which included a topic deviation. When examining which influence the attributions had on this behavior, the model was not significant ( $\chi^2(3) = 2.04, p = .565$ , Nagelkerke's  $R^2 = .02$ ). Thus, whether participants addressed the discussion topic was not influenced by attributions of discussion intentions, ( $b = .02, SE = .04, Odds = 1.02, p = .716$ ), the attribution of tolerable political views ( $b = -.06, SE = .05, Odds = 0.94, p = .219$ ), or the attribution of discussion skills ( $b = -.00, SE = .05, Odds = 1.00, p = .972$ ). When investigating the influence of these attributions on participants likelihood of criticizing, the overall model was significant ( $\chi^2(3) =$

43.63,  $p < .001$ , Nagelkerke's  $R^2 = .25$ ). It was revealed that the attribution of discussion intentions did not affect participants' likelihood of expressing critique ( $b = -.00$ ,  $SE = .05$ , Odds = 1.00,  $p = .972$ ). However, the attributions of more tolerable political views ( $b = -.07$ ,  $SE = .03$ , Odds = 0.93,  $p = .023$ ) and more discussion skills ( $b = -.07$ ,  $SE = .05$ , Odds = 0.93,  $p = .028$ ) decreased the likelihood of criticizing the preceding commenter. However, the effects were small.

Concerning participants' own uncivil behavior, it was found that individuals derailed the most from the discussion when exposed to a topic deviation. Therefore, in the first logistic regression, it was tested whether participants' attributions had an influence on this behavior. However, the model was not significant ( $\chi^2(3) = .08$ ,  $p = .151$ , Nagelkerke's  $R^2 = .07$ ). Neither the attribution of discussion intentions ( $b = .05$ ,  $SE = .06$ , Odds = 1.05,  $p = .424$ ), the attribution of tolerable political views ( $b = .06$ ,  $SE = .07$ , Odds = 1.06,  $p = .368$ ), nor the attribution of discussion skills ( $b = .02$ ,  $SE = .07$ , Odds = 1.02,  $p = .788$ ) influenced participants' derailing behavior. Furthermore, it was found that people mostly violated democratic norms when exposed to such norm violations before. Therefore, next it was investigated whether participants' attributions of the preceding commenter influenced the likelihood of such norm violations. Again, the model was not significant ( $\chi^2(3) = 1.15$ ,  $p = .764$ , Nagelkerke's  $R^2 = .01$ ). Consequently, participants' violations of democratic norms were not influenced by attributions of discussion intentions ( $b = .01$ ,  $SE = .06$ , Odds = 1.01,  $p = .906$ ), the attribution of tolerable political views ( $b = -.04$ ,  $SE = .06$ , Odds = 0.96,  $p = .441$ ), or the attribution of discussion skills ( $b = .05$ ,  $SE = .06$ , Odds = 1.05,  $p = .410$ ). Thus, individuals' uncivil behavior was not affected by the attributions.

#### **4. Discussion**

The main purpose of the current study was to explore how distinct types of incivility differ regarding recipients' attributions of the sender and how their behavioral reactions alter. It was proposed that, by this, new ways can be found to categorize incivility based on its

effects and not simply its manifestation. In line with assumptions about attribution processes (Malle, 1999) and cooperative communication (Allwood et al., 2000; Grice, 1975), we expected that all types of incivility decrease recipients' attributions about senders' intentions to discuss, tolerable political views, and discussion skills (H1.1a – H1.5c). Indeed, for most types of incivility, these assumptions were confirmed. Comments including ironic sarcasm or disrespect towards other discussants led to lower attributions about all the three dimensions. This indicates that these types of incivility signal a lack of cooperation on a rather universal level. However, comments including a topic deviation only decreased attributions of discussion intentions, but not tolerable political views or discussion skills. Nevertheless, as shared intentions are a rather basic precondition for cooperative behavior (Allwood et al., 2000; Tomasello et al., 2005), this is probably sufficient to intrude discussions. Here, Hopp (2019) argued that intentionally suppressing or limiting a deliberative discussion is inherently uncivil as it disrupts the democratic exchange. Accordingly, on a perceptual level, people seemingly assume that the people who deviate from the topic in a discussion are less willing to cooperate within this exchange. Such intentional refusals to cooperate are perceived as worse than uncooperative behavior due to ignorance (Kachel et al., 2018). Nevertheless, it is surprising that individuals did not attribute lower levels of tolerable political views or discussion skills to persons who deviate from the discussion topic compared to civil comments. One would assume that the violation of basic cooperative norms (Grice, 1975) would also decrease such attributions. However, it might be that this type of incivility does not provide sufficient information about the other attribution dimensions as it focuses on discussion suppression rather than providing arguments or own perspectives. Moreover, compared to civil comments, violations of democratic norms did not lead to decreased attributions of discussion skills. This indicates that people sharply distinguish violators' discussion skills from their intent and their views. Thus, it can be assumed that people consider this behavior to be performed quite deliberately.

Moreover, although it was already found that individuals assess misleading information behavior as uncivil (Stryker et al., 2016), we could not find that the exposure to misleading information led to decreased attributions about senders' intentions to discuss, tolerable political views, and discussion skills. Identifying misleading information is a complex process that is biased in many ways (see Lewandowsky et al., 2012 for an overview). Indeed, as indicated by the manipulation check, people had the most problems identifying misleading information. Nevertheless, the manipulation check indicated that most participants recognized the norm violations within the comment, including misleading information. Hence, misleading information is seemingly not affected by the dimensions of cooperative communication that we assessed.

#### *4.1. Differences between distinct types of incivility*

We further hypothesized that there are differences between the types of incivility regarding recipients' attributions (H2a – c) and asked how these differences look like (RQ1). Supporting H2a – c, we found differences regarding all three attributional levels based on the type of incivility – albeit, not all types of incivility differed significantly from each other.

**5.1.1. Discussion intentions.** We found that of all types of incivility, disrespect towards others decreased attributions of discussion intentions by far the most. Even individuals, who did not address the discussion topic, were attributed with more discussion intentions. This outcome is rather surprising as, in this condition, the comment included the same arguments that were included in the other conditions that did not deviate from the topic. However, Kluck and Krämer (2021) already demonstrated that disrespect leads to increased attributions of aggressive intent. Therefore, disrespectful behavior towards others might be interpreted as the ultimate signal of a lack of cooperation as it induces conflict. Topic deviation, ironic sarcasm, and violations of democratic norms did not differ from each other regarding attributions of discussion intentions, but the mean scores were at a moderate level. As these types of incivility are quite diverse, this is a surprising finding. Of course, people are

likely to infer a different set of motives from different behaviors. However, discussion intentions seem to be perceived on a similar level. Therefore, regarding individuals' intentions to cooperate, these types of incivility can be clustered as people assess them on a similar level. Noteworthy, although the attributions of discussion intentions did not differ between the civil condition and the condition including misleading information, the exposure to misleading information did not lead to significantly more attributions of discussion intentions than the exposure to ironic sarcasm or violations of democratic norms.

**5.1.2. Tolerable political views.** We also found that there were differences between the distinct types of incivility in terms of attribution of tolerable political views. First, the lowest ratings of tolerable political views were given to those that violate democratic norms. This is not surprising as the violation against democratic norms often roots in violators' beliefs (e.g., towards social groups or democracy itself; Papacharissi, 2004; Rossini, 2020). However, what is rather surprising is that disrespect towards other discussants did not differ significantly from violations of democratic norms regarding the attribution of tolerable political views. Although types of incivility related to these norm violations are most often distinguished in literature (Muddiman, 2017; Papacharissi, 2004; Rossini, 2020), they might share some similarities on a perceptual level. Since our analysis revealed that disrespect is associated with the lowest levels of discussion intentions, people seemingly do not regard disrespect towards others as "heated disagreement" (Papacharissi, 2004, p. 262). Rather, it can be assumed that people regard this type of incivility as disrespectful beyond individual attacks. When individuals communicate in this way, recipients are probably less convinced of the sender's ethical considerations, which, Allwood and colleagues (2000) argue, is a prerequisite for cooperative communication. The result is all the more astonishing as in the condition of disrespect, no additional political views were included, but only personal attacks on the preceding commenter.

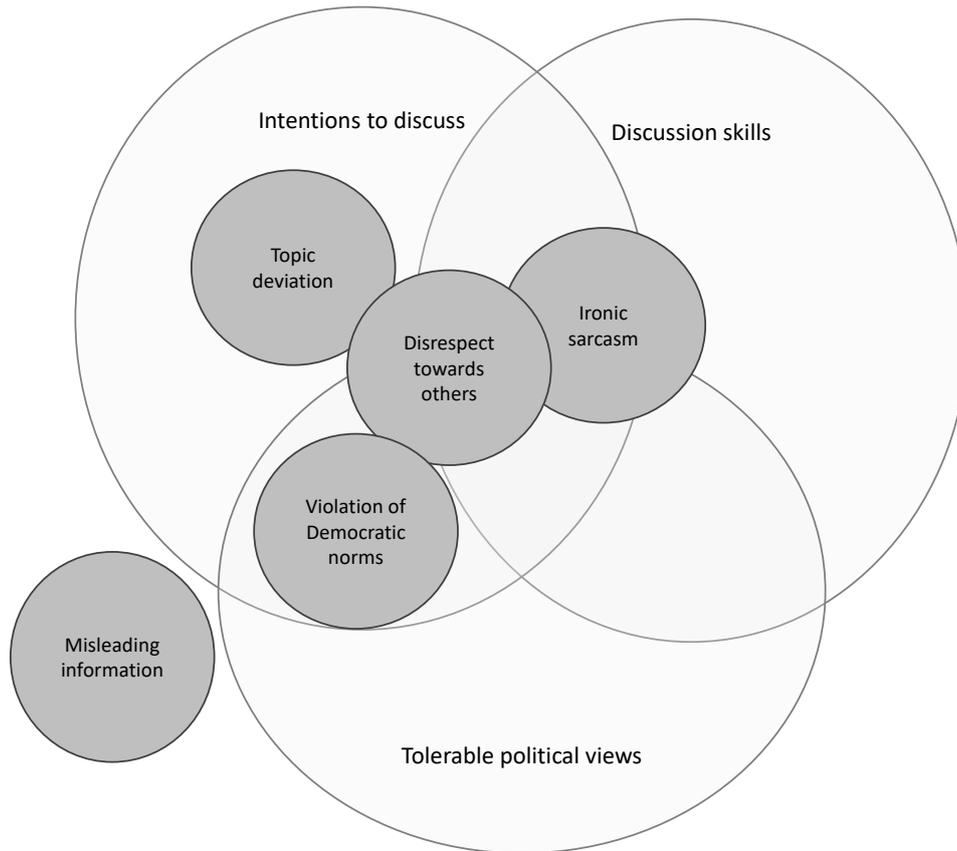
Furthermore, we found that the differences in attributions about tolerable political views were significant between the conditions of ironic sarcasm and violations of democratic norms, but not between the conditions of ironic sarcasm and disrespect towards other discussants. Again, in the condition of ironic sarcasm, the uncivil sender did not include a dissenting or additional view. However, as ironic sarcasm usually obscures people's real stance, this might also indicate less trustworthiness in an unknown sender's ethical considerations, as a common ground for the interpretation of the message is missing (Anderson & Huntington, 2017). This demonstrates that people do not simply make inferences about others' political views by the content of their statements but also by how they communicate. Interestingly, ironic sarcasm did not differ from topic deviation and misleading information, which both in turn did not differ significantly from the civil condition. Thus, disrespect towards others and violations of democratic norms lead to the attribution of less tolerable political views, whereas ironic sarcasm seems to be associated with more tolerable political views – probably because senders' opinion is more difficult to assess. Moreover, people might rather consider such behavior accusatory instead of opinionated (Stroud & Muddiman, 2013).

**5.1.3. Discussion skills.** Finally, we also found differences between the distinct types of incivility regarding the attribution of discussion skills. Participants attributed the fewest skills to senders, who showed disrespect towards other discussants. This was followed by ironic sarcasm, which did not differ significantly from the condition of disrespect. Thus, these kinds of incivility seem to decrease trust in senders' ability to cooperate. Probably, recipients assume that senders of such messages are not able to communicate their opinion appropriately. However, the reasons might differ. As disrespect is likely to be associated with increased aggressive motives (Kluck & Krämer, 2021), people might rather assume reduced self-control, whereas ironic sarcasm might rather be associated with an inability to communicate one's thoughts clearly and comprehensibly (see also Grice, 1975). However,

ironic sarcasm did not differ from violations of democratic norms or topic deviation, which again did not differ from the civil condition. This indicates that usage of ironic sarcasm is perceived as moderately uncooperative regarding all three dimensions of cooperation tested in our study. Violation of democratic norms, topic deviation, and misleading information did not differ significantly from each other and were not assessed differently than civil comments. Thus, these types of incivility are apparently not associated with a lack of discussion skills – which is rather surprising. However, this underscores that different types of incivility operate in very different ways.

Taken together, we found that disrespect consistently resulted in low ratings among all attribution levels. This result can explain why strong markers of disrespect towards others are assessed as more uncivil than other forms of incivility (Kenski et al., 2020; Muddiman, 2017). As this type of incivility negatively affects the basic requirements for cooperation as proposed by Allwood et al. (2000), people assumably perceive an overarching lack of cooperativeness. In contrast, violations of democratic norms especially decreased attributions of tolerable political views. Thus, in accordance with Muddiman's (2017) finding that personal-level incivility is perceived as more uncivil than public-level incivility, people seem to regard this type of incivility to be rather cooperative than disrespectful towards others as they infer more intentions to discuss and higher discussion skills from this behavior.

Moreover, we also found that the usage of more subtle types of incivility can have a negative effect on recipients' attributions about their sender. For instance, topic deviation was associated with a lack of discussion intent. Moreover, ironic sarcasm led to low attributions of discussion skills. As outlined in *Figure 1*, this allows us to categorize incivility based on the attributions people make.



**Figure 1.** Visual categorization of the distinct type of incivility based on attributions of cooperation.

#### 4.2. Behavioral outcomes

We further argued that the behavioral outcomes provide an additional way to identify differences between the distinct types of incivility. Therefore, we investigated which impact the exposure of the different uncivil messages has on participants' behavior. First, we examined whether distinct types of incivility trigger different uncivil reactions (RQ2). However, only a few cases of uncivil behavior emerged at all. We found that the exposure to a topic deviation more often led to further derailing than the other conditions. Likewise, we found that democratic norms were most often violated when people were exposed to this type of incivility before. Nevertheless, the number of cases was very small. We assume that due to participants' unfamiliarity with our discussion platform, the communicated rules, and the civil behavior of the other discussants, people rather followed a civil norm. Even when explicitly addressing a norm violation, participants often reacted in a respectful and understanding

manner. Moreover, previous research shows that uncivil commenting behavior is rare among online commenters. The majority seem to write such messages in more exceptional cases to counteract previous incivility (Kluck & Krämer, 2020; Moor et al., 2010). Accordingly, in our study, findings that preceding incivility triggers further uncivil behavior or identifies specific patterns of uncivil behavior among the distinct types of incivility were limited.

We also wanted to know how the distinct types of incivility would affect participants' discussion behavior beyond incivility. We found that regardless of the experimental conditions, most participants wrote a comment (RQ3a&b), and our content analyses revealed that most comments were rather topically related (RQ3c) and argumentative (RQ3d). However, the analyses showed that there were differences between the conditions in addressing the topic of the discussion. Descriptive values indicated that topic deviations, in particular, have led to people deviating from the topic of discussion. This is not very surprising as the rules of our discussion platform stated that people should explicitly address the preceding comment. However, even here, most participants managed to address both the new topic introduced in the condition of topic deviation and the actual discussion topic. Thus, it is rather remarkable how strongly people have adhered to all the rules of the discussion forum. We additionally investigated how often the distinct types of incivility were criticized by recipients. Here, we have found clearer results. It turned out that people criticized the senders of uncivil messages mainly when they faced disrespect, followed by violations of democratic norms. This corresponds to the findings that disrespect was consistently assessed low on our attributional dimensions of cooperative behavior, and violation of democratic norms was associated with a low extent of tolerable political views.

Furthermore, regarding RQ4, we did not find that participants' attributions influenced derailing behavior or violations of democratic norms. Likewise, the attribution variables did not influence whether people address the discussion topic or not. This is comprehensible as these behaviors are less addressing the sender of the message but rather refer to thematic

aspects. However, we found that critique behavior was more likely when people attributed lower levels of tolerable political views and lower levels of discussion skills. Attributions about the commenter's discussion intentions did not affect critique behavior. In this context, Kachel et al. (2018) found that humans tend to emotionally protest against intentional uncooperative behavior and to teach individuals who behave uncooperatively due to ignorance. Indeed, when expressing critique, participants often engaged in teaching behavior by pointing out rules and explaining why certain behavior is destructive. People are seemingly more likely to engage in such behavior when they perceive increased cooperative potential due to personal characteristics and not their intentions. Therefore, critique is probably a mere form of punishment or an expression of aversion as proposed by (Gervais, 2015). It is important to note that due to participants' laudable discussion behavior, the validity of our inferences in terms of behavioral outcomes is very limited. Nevertheless, we were able to identify some tendencies that deserve further attention in future studies.

#### *4.3. Theoretical and practical implications*

Our study yielded some important theoretical and practical implications. First, our approach provides a more nuanced way to categorize incivility on a perceptual level. So far, differences between distinct types of incivility on a perceptual level have only been found by evaluating comments (Kenski et al., 2020) or descriptions of certain discussion behaviors (Stryker et al., 2016) in terms of the perceived extent of incivility. Extending these findings, we demonstrated that people perceive differences in the person-related causes for uncivil behavior. Thus, if recipients of uncivil messages make differences regarding the origins of the behavior, research should do this, too. This measure of incivility allows a more reliable assessment of perceived incivility as it goes beyond the mere appraisal of a message. In this way, we provide a new theoretical and methodological approach for categorizing incivility. Although in our study effects on behavioral variables were limited, from a theoretical

perspective, attributions about others' willingness and ability to cooperate allow new ways to understand the consequences of incivility.

In this way, we can better tailor interventions against different types of incivility. For instance, in the case of topic deviation, it might be more helpful to encourage individuals to respond to such comments by teaching them, that people who deviate from the discussion topic might still have an interest to discuss the issue when seeing connections between the topics or understanding the relevance of the initial discussion topic. In contrast, in the case of ironic sarcasm, it might be more helpful to encourage individuals to react to the comment by explaining that a brief request to refrain from such behavior can improve the person's discussion contributions. The other way around, when addressing uncivil commenters, it might be more fertile to explain to them which attributions people make about them rather than to tell them that their comments are uncivil. For example, in the case of comments including disrespect towards others, telling the sender that others perceive them as unwilling to discuss, unskilled, and associate intolerable views with them could be more effective than explaining to them that the message is inappropriate. However, further research is needed to develop and evaluate respective interventions.

#### *4.4. Limitation*

Some limitations of the current study must be noted. First, although the sample is relatively heterogeneous, it is not representative. Therefore, the generalizability of our results is limited. Regarding the external validity, the study has some shortcomings but also provides some strengths. On the one hand, we developed a discussion forum with rather unusual rules to control the discussion and induce a cooperative situation. Thus, the usage was not only unfamiliar to the participants but discussion norms were made very salient, which might have influenced participants' discussion behavior more than it was intended. On the other hand, we achieved that the participants – for the most part – felt that the discussion was real and that they were talking to real people. Another limitation is that we only included a limited number

of types of incivility. However, there are many more types of incivility that can manifest in various ways. Further limitations result from our measurements. First, as no standardized measures that would have been suitable for our study existed, we drew on self-developed scales. Therefore, the validity of the scales can only be determined to a limited extent. Second, we only examined a few reactive behaviors. Other behavioral outcomes should be addressed in future studies.

## **5. Conclusion**

This study provides a new fundament to measure and categorize different types of incivility. Drawing on assumptions about cooperative behavior and communication norms (Allwood et al., 2000; Bormann et al., 2021; Grice, 1975; Tomasello et al., 2005), we found that people make different attributions about senders of uncivil messages depending on the type of incivility. This result highlights that incivility is not a message characteristic but a social act that is processed by recipients accordingly. Therefore, future studies examining the effects of incivility can draw on this framework to identify mechanisms in which incivility works. Likewise, interventions can be designed in a more tailored way by addressing the perceived person-related causes of uncivil commenting behavior.

## References

- Allwood, J., Traum, D., & Jokinen, K. (2000). Cooperation, dialogue and ethics. *International Journal of Human-Computer Studies*, 53(6), 871–914.  
<https://doi.org/10.1006/ijhc.2000.0425>
- Anderson, A. A., & Huntington, H. E. (2017). Social Media, Science, and Attack Discourse: How Twitter Discussions of Climate Change Use Sarcasm and Incivility. *Science Communication*, 39(5), 598–620. <https://doi.org/10.1177/1075547017735113>
- 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*. Advance online publication.  
<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*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-56273-5>
- 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 + Society*, 5(3).  
<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>
- 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>
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (3rd ed.). SAGE.
- Gervais, B. T. (2015). Incivility Online: Affective and Behavioral Reactions to Uncivil Political Posts in a Web-based Experiment. *Journal of Information Technology and*

- 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, H. P. (1975). Logic and Conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax And Semantics* (3rd ed., Issue January, pp. 41–58). Academic Press.
- 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>
- Han, S.-H., Brazeal, L. M., & Pennington, N. (2018). Is Civility Contagious? Examining the Impact of Modeling in Online Political Discussions. *Social Media + Society*, 4(3). <https://doi.org/10.1177/2056305118793404>
- Hopp, T. (2019). A network analysis of political incivility dimensions. *Communication and the Public*, 4(3), 204–223. <https://doi.org/10.1177/2057047319877278>
- Hutchens, M. J., Silva, D. E., Hmielowski, J. D., & Cicchirillo, V. J. (2019). What’s in a username? Civility, group identification, and norms. *Journal of Information Technology & Politics*, 16(3), 203–218. <https://doi.org/10.1080/19331681.2019.1633983>
- Kachel, U., Svetlova, M., & Tomasello, M. (2018). Three-Year-Olds’ Reactions to a Partner’s Failure to Perform Her Role in a Joint Commitment. *Child Development*, 89(5), 1691–1703. <https://doi.org/10.1111/cdev.12816>
- 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>
- 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*. Advanced Online Publication. <https://doi.org/10.1177/14614448211000666>
- Kluck, J. P., & Krämer, N. C. (2020). It’s the aggression, stupid! An examination of commenters’ motives for using incivility in online political discussions. *International Conference on Social Media and Society*, 164–173. <https://doi.org/10.1145/3400806.3400826>

- 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>
- Malle, B. F. (1999). How People Explain Behaviour: A New Theoretical Framework. *Personality and Social Psychology Review*, 3(1), 23–48. [https://doi.org/10.1207/s15327957pspr0301\\_2](https://doi.org/10.1207/s15327957pspr0301_2)
- Malle, B. F. (2011). Attribution Theories: How people make sense of behavior. *Theories in Social Psychology*, 72–95.
- Mooney, A. (2004). Co-operation, violations and making sense. *Journal of Pragmatics*, 36(5), 899–920. <https://doi.org/10.1016/j.pragma.2003.10.006>
- Moor, P. J., Heuvelman, A., & Verleur, R. (2010). Flaming on YouTube. *Computers in Human Behavior*, 26(6), 1536–1546. <https://doi.org/10.1016/j.chb.2010.05.023>
- Muddiman, A. (2017). Personal and public levels of political incivility. *International Journal of Communication*, 11, 3182–3202.
- 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>
- Rösner, L., & Krämer, N. C. (2016). Verbal Venting in the Social Web: Effects of Anonymity and Group Norms on Aggressive Language Use in Online Comments. *Social Media + Society*, 2(3). <https://doi.org/10.1177/2056305116664220>
- 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. (2019). Disentangling Uncivil and Intolerant Discourse. In R. G. Boatright, T. Shaffer, S. Sobieraj, & D. Young (Eds.), *A Crisis of Civility? Contemporary Research on Civility, Incivility, and Political Discourse*. (pp. 142–158). Routledge.
- Rossini, P. (2020). Beyond Incivility: Understanding Patterns of Uncivil and Intolerant Discourse in Online Political Talk. *Communication Research*. <https://doi.org/10.1177/0093650220921314>
- Stroud, N. J., & Muddiman, A. (2013). Selective exposure, tolerance, and satirical news.

*International Journal of Public Opinion Research*, 25(3), 271–290.

<https://doi.org/10.1093/ijpor/edt013>

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>

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., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *Behavioral and Brain Sciences*, 28(5), 675–691; discussion 691-735. <https://doi.org/10.1017/S0140525X05000129>

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/10.1177/0093650216671854>

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. *Studies in Communication / Media*, 7(4), 525–554. <https://doi.org/10.5771/2192-4007-2018-4-525>

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>

## Appendix A: Pretest

To develop the stimulus material, we first constructed a civil comment as a baseline for each topic (i.e., the expansion of referendums in Germany or the introduction of an unconditional basic income in Germany). Then, within an iterative process, the uncivil comments were constructed for each topic. The concrete manifestation of distinct types of incivility has been differed between the different topics to increase the generalizability of the results. Care has been taken to ensure that the comments' core arguments remain comparable (except for topic deviation) and that there is no severe overlap between different types of incivility. Thus, in sum, twelve comments have been developed. To enhance the quality of the material, it was pretested ( $N = 47$ ,  $M_{\text{age}} = 42.60$ ,  $SD = 13.70$ , 68.1% female) to identify the remaining flaws of the comments. Drawing on a within-subjects design, all participants rated eight items for each comment. The items were worded straightforwardly and asked whether features corresponding to norm violations were present (7-point Likert scale from 1 = *do not agree at all* to 7 = *totally agree*). Comments were presented in random order and without any preceding comments. Results indicated that most norm violations were assessed as intended (see *Table A1* and *Table A2*). However, as some comments were rated high on items that did not address the respective norm violations, we refined the comments to avoid undesired perceptions. For instance, comments including violations of democratic norms were more strongly marked as an opinion, and the reference to the topic was emphasized more strongly. Deeper conversations with some participants helped us to identify problematic formulations. Noteworthy, these conversations revealed that participants largely perceived the norm violation as it was intended but were unsure about some items and did not want to "misjudge" a comment. For instance, they could conceptually differentiate between misleading information and the violation of democratic norms but were unsure whether stereotyping can be considered a form of doubtful fact.

Table A1

*The mean values of the ratings of the eight items for comments addressing the topic of the expansion of referendums in Germany.*

The comment...	Topic: Referendums											
	Civil comment		Misleading information		Ironic sarcasm		Topic deviation		Disrespect to others		Democratic norm violation	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
...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 statements	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 strongly 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

Table A2

*The mean values of the ratings of the eight items for comments addressing the topic an unconditional basic income in Germany.*

The comment...	Topic: Basic income											
	Civil comment		Misleading information		Ironic sarcasm		Topic deviation		Disrespect to others		Democratic norm violation	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
...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 statements	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 strongly 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

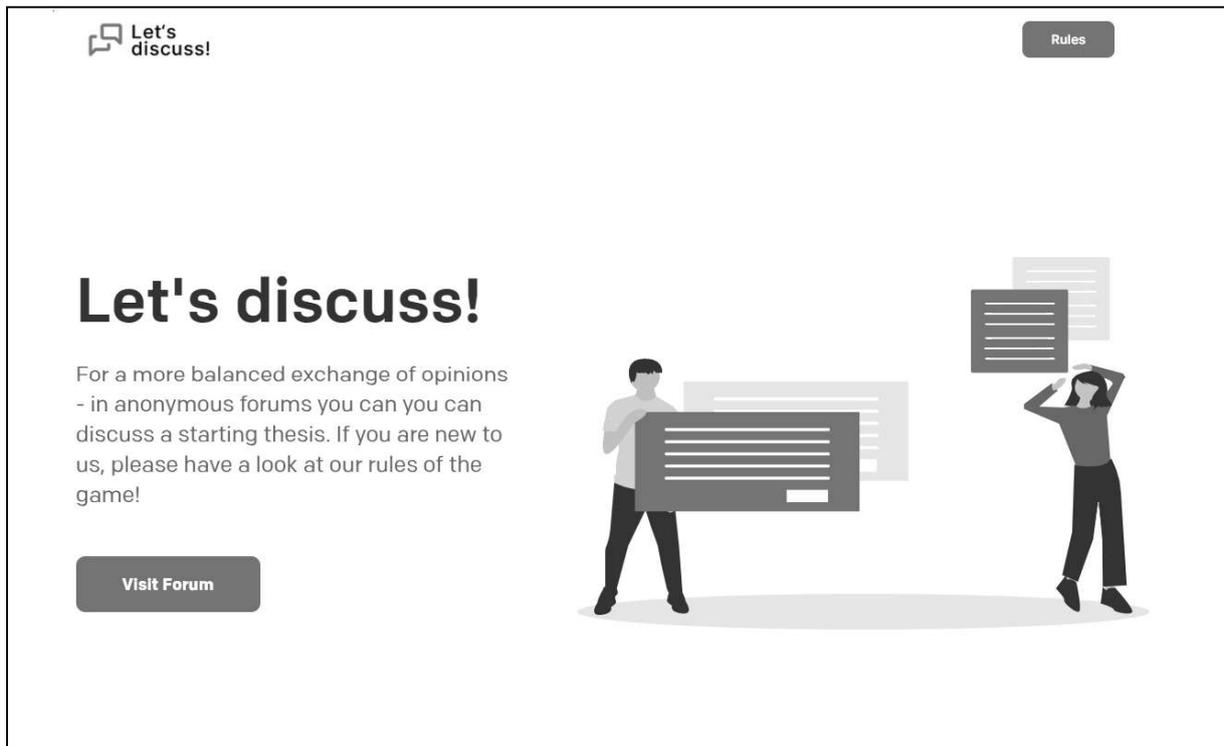
## Appendix B: Materials

Before being directed to the discussion platform, participants were presented with the forum's rules within the questionnaire. At the beginning of the rules, the participants were informed that they would be automatically assigned to a discussion topic and that the opinion climate of the discussion forum is balanced by the following topic-related questions. The following is an abbreviated overview of the rules presented to the respondent:

- The first rule is that discussions on "Let's Discuss!" are turn-based
  - For each round, all participants (maximum six people) write a comment one after the other
  - The order in which the participants write their comments will be determined randomly at the beginning of each round.
- Reply only to the previous comment
  - Only the person who writes the first comment at the beginning of the discussion can express his opinion on the topic once independently of other comments.
  - Even if another comment seems more interesting at the first moment, you are only allowed to react to the previous comment.
  - There should always be a reference to the previous comment - but of course, new arguments can be introduced.
  - If you cannot or do not want to reply to the previous comment, you can skip a round by clicking "Pass." In this case, the next person must reply to the corresponding comment.

At the welcome page of the discussion forum, participants had the opportunity to look at the forum's rules again or to join a discussion (see Figure B1). When they chose to join a

discussion, they were asked about their opinions on various societal issues to suggest that they would be automatically assigned to a discussion based on these answers (see Figure B2). However, subjects were equally allocated to one out of two fixed topics (i.e., the expansion of referendums in Germany or the introduction of an unconditional basic income in Germany). Please note that materials are provided for the topic of referendums in Germany.



**Figure B1.** Welcome page of the discussion forum.

## What Is Your Opinion?

To keep the climate of opinion in the discussions as balanced as possible, we need some information about your opinion on some topics.  
You will then be assigned to a forum based on your answers.

Do you think that people in Germany have enough opportunities to directly contribute their voice to political issues (for example, in referendums)?

Yes  
 No  
 Undecided

Do you think that the unconditional basic income (i.e. a financial allowance from the state without anything in return) 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 B2.** Alleged questions about the opinion of the participants for the assignment of a discussion forum.

The participants moved to the actual discussion part (see Figure B3). First, information about the assigned topic was displayed, and the respondent is given time to read through it carefully. After predefined time intervals, a notification about the writing activity of another user is displayed, and after another time interval, his corresponding answer is displayed. In the design, special attention was paid to defining the times close to reality to give the participant the impression that he receives answers from a real person. The time intervals used can be seen in the following table (see Table B1).

**Table B1**

*Time Intervals of the Writing Notices and Responses in the Forum.*

Element	Writing 1	Answer 1	Writing 2	Answer 2	Writing 3	Answer 3
Time in Seconds	15	22	17	25	14	23,5

A total of one topic and three responses are displayed to each participant. Posts one and two are static per topic (see Figure B3), and post three was manipulated in accordance with the allocated condition (see Table B2). All posts were related to the assigned topic. After all, the content was displayed, and the subject is offered a choice of three different actions. The respondent can either post a reply to the previous comment, pass one round or leave the discussion entirely. After submitting the answers, deciding to pass one round or leave the discussion, the respondent is returned to the questionnaire.

# Discuss!

Discuss with other people on the topic given:

## Democracy

There is always a debate about whether citizens should have more of a say in major 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 directly on whether they are for or against a particular issue. Many would like to see an expansion of these votes - including at the federal level. Others believe that complex political issues cannot be broken down into yes and no questions. From a scientific perspective, it is difficult to determine whether direct forms such as referendums achieve a better democratic result than decisions made by elected representatives. Studies cannot provide any reliable statements on this.

What do you think? What are the arguments for or against expanding forms of direct democracy?



## First message

Written by JR89

Hello everyone, then I'll make the beginning. I am still unsure, I think direct democracy has advantages and disadvantages. Although the people then have more influence on certain political decisions but on the other hand, the various parties can manipulate the mood in their interest. What do you think?



## Reply to JR89

Written by Soly

Thanks for the start JR89... I think the problem is that citizens don't always have the expertise to properly weigh the decision. Therefore, politicians may have the better overview of what is best for everyone.



## Reply to Soly

Written by Pat

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 casts a vote directly for themselves...

Now it's your turn!

Reply to Pat

Pass one Round

Leave Discussion

Figure B3. Discussion at participants' turn.

**Table B2***Manipulation of the comments.*

Type of incivility	Comment
Civil	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...
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 provide citizens the necessary information without any problems. In the end, this is the only true representation of interests for each individual...
Ironic 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...
Disrespect to other discussants	@Soly, I think this is a really stupid argument from you, apparently, your mind is somewhat limited. Still, you can give citizens the necessary information through information campaigns. I think it sucks anyway if not everyone with his voice directly can represent their own interests ...
Violation of democratic norms	Hey Soly, you can also give citizens the necessary information through 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 best kicked directly out of the parliaments if we can represent our interests directly with our vote...

*Note.* Comments are only presented for the topic of referendums. In the original German version, comments included minor misspellings to increase the authenticity of the comments.

### **Appendix C: Validation of measures**

To validate the self-constructed measures, we additionally measured the perception of persons' *warmth* (e.g., sincere or good-natured) *and competence* (e.g., intelligent or confident) as proposed by Fiske and colleagues (1999). Likewise, we measured the *attribution of aggressive commenting motives* (e.g., to humiliate others in the discussion ) adopted from Kluck and Krämer (2021). Again, both scales were measured on a 7-point Likert scale (from 1 = *do not agree at all* to 7 = *totally agree*). We expected that there would be a strong positive between the warmth factor and the attribution of tolerable political views. Likewise, we expected a strong positive correlation between the competence dimension and the attribution of discussion skills. Additionally, we assumed that there is a strong negative correlation between the attribution of discussion intentions and the attributions of aggressive motives.

As can be seen in Table C1, the scales correlated as expected, respectively. However, it must be noted that all scales are highly correlated with each other. Foremost, discussion intentions correlated strongly with the warmth and competence dimensions. This pattern might emerge because people draw inferences about more stable person characteristics from the attributed behavioral motives (Reeder, 2009). However, as the items for the measures of attribution have been formulated in a straight manner and correlated with the other scales as intended, we considered them to be suitable for the study's purpose. Since the measures of attribution were also relatively highly correlated, we further checked for indices of multicollinearity. However, we could not find any evidence for multicollinearity.

**Table C1***Correlations between the dependent measures and the scales for validation.*

---

	(1)	(2)	(3)	(4)	(5)
(1) Discussion intentions					
(2) Tolerable political views	.604*				
(3) Discussion skills	.695*	.608*			
(4) Competence	.660*	.501*	.625*		
(5) Warmth	.745*	.653*	.651*	.715*	
(6) Aggressive motives	-.575*	-.494*	-.541*	-.433*	-.601*

---

\*p &lt; .001

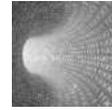
**References**

- Fiske, S. T., Xu, J., Cuddy, A. C., & Glick, P. (1999). (Dis) respecting versus (Dis)liking : Status and Interdependence Predict Ambivalent Stereotypes of Competence and Warmth. *Journal of Social Issues*, 55(3), 473–489.
- 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>
- Reeder, G. D. (2009). Mindreading: Judgments About Intentionality and Motives in Dispositional Inference. *Psychological Inquiry*, 20(1), 1–18. <https://doi.org/10.1080/10478400802615744>

## Article III

The following article is reused from:

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>



Article

# “What an idiot!” – How the appraisal of the writer of an uncivil comment impacts discussion behavior

new media & society

1–20

© The Author(s) 2021



Article reuse guidelines:

[sagepub.com/journals-permissions](http://sagepub.com/journals-permissions)

DOI: 10.1177/14614448211000666

[journals.sagepub.com/home/nms](http://journals.sagepub.com/home/nms)



Jan P Kluck<sup>ID</sup> and Nicole C Krämer

University of Duisburg-Essen, Germany

## Abstract

Despite incivility in online discussions being linked to various negative effects, less is known about the mechanisms of how incivility works. So far, explanations by social perception have been neglected. Therefore, drawing on the multiple inference model, this study employs an attribution theoretical approach to examine whether the motives and traits that people attribute to senders of uncivil or opposing comments affect their intentions to join a discussion. Employing a 2 (incivility vs. no incivility) × 2 (like-minded vs. opposing stance) between-subjects design, data from an online experiment ( $N=452$ ) were analyzed applying a path model (SPSS AMOS). Results revealed that participants attributed more aggressive and less nonaggressive motives to senders of uncivil messages. The attribution of aggressive motives consequently increased hostile inferences about the target. A similar pattern occurred when individuals were exposed to an opposing stance. In result, hostile inferences about the sender's traits decreased participants' willingness to discuss.

## Keywords

Attribution process, incivility, online participation, online political discussions, self-categorization, user comments

## Introduction

Exchanging opinions and ideas in a respectful and constructive manner is an essential part of our democracy (Herbst, 2010). Although the Internet has the potential to foster the public discourse by connecting citizens with different views, scholars noticed that the

### Corresponding author:

Jan P Kluck, Social Psychology, Media and Communication, University of Duisburg-Essen, Forsthausweg 2, 47057 Duisburg, Germany.

Email: [jan.kluck@uni-due.de](mailto:jan.kluck@uni-due.de)

online political exchange has become increasingly uncivil (e.g. Coe et al., 2014). Therefore, incivility in online political discussions has received increasing scientific attention. Scholarship has demonstrated that the exposure to incivility fosters negative emotions, increases disagreement, and decreases open-mindedness toward opposing stances (Chen and Lu, 2017; Hwang et al., 2018). Moreover, incivility was found to discourage users from participating in online discussions (Han and Brazeal, 2015) and, consequently, threatens diversity in political online debates. Although this is a very serious effect, little is known about the underlying mechanisms of how incivility discourages individuals from joining discussions. While there is scholarly consensus on the subjective nature of incivility (Chen et al., 2019), it is usually treated as a distinct attribute of messages or discussions. This perspective, however, overshadows that social perception directs how people explain others' behavior. Considering that online discussions are characterized by the social exchange between citizens, we assume that people do not refrain from discussing with others because they simply see an uncivil message. Rather, we assume that the negative assessment about the sender of an uncivil message drives this decision. In other words, people do not want to discuss with others they consider to be "idiots." The question is however, how do people come to such an assessment?

In this vein, attribution theories are helpful to understand how people assess incivility from an interpersonal perspective. Put broadly, these theories describe how individuals explain others' behavior by making casual attributions and/or by forming impressions about the actors. Based on these assumptions, the *multiple inference model* (Reeder, 2009; Reeder et al., 2002) was developed to approach attribution processes from the perspective of social perception. The model suggests that perceivers make inferences about traits and motives within others to understand their reasons for acting. These judgments can determine whether people classify the observed action as hostile and drive their reaction (Anderson and Bushman, 2002). Thus, the attributions recipients make about senders of uncivil messages is important because they influence whether people expect this behavior to persist.

However, it is unlikely that people will be assessed solely on the wording they use in their comments. In online discussions, individuals are often confronted with opposing views. Research indicates that the stances represented in discussion comments influences how people feel about and react to uncivil messages (Chen and Lu, 2017; Gervais, 2017). In this respect, the *self-categorization theory* (Turner et al., 1987) states that people favor individuals from social groups with whom they share certain similarities and discriminate against individuals with whom they do not identify. This basic human tendency has been shown to bias attribution processes in favor of like-minded individuals (Reeder et al., 2005). Hence, we expect that the sender's stance represented in a comment is of importance for how messages are construed by the recipients.

Based on the assumptions of the multiple inference model (Reeder, 2009), we conducted an online experiment to investigate whether individuals attribute more aggressive motives and hostile traits to a sender of an uncivil message compared with a sender of a civil message. Following the self-categorization theory (Turner et al., 1987), we further examined the impact of the comment's represented stance on this process. We also investigated how the resulting assessment influences recipients' willingness to join the discussion. As social perception is affected by individuals' own attributes, the influence of

perceivers' personality was examined, too. In doing so, we focused on peoples' "dark side" of personality since incivility is considered as anti-normative behavior.

## Literature review

### *Incivility in online public discourse*

On a basic level, incivility has been conceptualized as the violation of social norms of interaction (Jamieson and Hardy, 2012). Therefore, research has often focused on the use of disrespectful or insulting language to conceptualize uncivil online behavior. For instance, Coe et al. (2014: 600) defined incivility as "features of discussion that convey an unnecessarily disrespectful tone toward the discussion forum, its participants, or its members." However, the concept of incivility is subjective by nature (Herbst, 2010). In this vein, Chen et al. (2019) suggested that incivility should be assessed by its outcomes rather than being based on the usage of particular words or phrases. However, the underlying mechanisms of how incivility affects individuals own behavior are not yet sufficiently researched.

Following the *general aggression model* (Anderson and Bushman, 2002), an integrative framework that explains aggressive behavior, humans develop different knowledge structures based on their experiences, including beliefs, attitudes, schemata, and behavioral scripts. These structures form an individual's personality and guide how social events are appraised. The model proposes that person characteristics and situation-based factors influence an individual's affect, cognitions, and feelings of arousal. These, in turn, impact how the individual appraises a social event by drawing inferences about the actor and the situational aspects. As a result, individuals decide on how to respond—that is, either in an aggressive way or not. In line with these assumptions, Rösner et al. (2016) found that uncivil comments can increase hostile cognitions. Other scholars showed that uncivil user discussions can affect feelings of anger, anxiety, or enthusiasm (Gervais, 2017; Hutchens et al., 2019). Moreover, it was demonstrated that incivility increases recipients' own use of incivility (Chen and Lu, 2017; Gervais, 2017; Ziegele et al., 2018), their willingness to participate politically online apart from the discussion (e.g. Chen and Lu, 2017), and the likelihood of expressing disagreement (Hwang et al., 2018). Even more, Han and Brazeal (2015) found that people were less likely to join an online discussion when it was uncivil.

Nevertheless, the processes that mediate the relationship between individuals' emotions and cognitions after exposure to incivility, as well as their behavioral intentions, remain largely unexplored. Here, the general aggression model provides a promising framework to explain the intersection between internal states and behavioral intentions.

### *Processing incivility from the perspective of attribution theories*

To describe how individuals appraise other people's behavior, attribution theories are usually accounted by social psychologists. These theories suggest that people rely on certain attributions about others' traits and/or situation-related aspects when they want to understand others' behavior. Reeder (2009), however, argues that behavior that is thought

to be under the actor's control is assessed differently from behavior that is thought to be performed against the actor's will (e.g. stumbling or sweating). He states that only unintentional behavior is assessed by making single inferences drawn from the situation or the actor. In terms of intentional behavior, he postulates a multiple inference model suggesting that perceivers also search for the actor's underlying motives under consideration of situational cues and combine them with trait inferences. In doing so, individuals obtain a comprehensive picture to explain others' behavior, helping them to appraise a social situation (Reeder, 2009). This being said, it is very unlikely that individuals think that others accidentally write user comments. Moreover, there is robust evidence that people have a general tendency to overestimate the influence of dispositional factors like personality traits or attitudes as explanation for others' behavior (i.e. fundamental attribution error; Ross, 1977). Therefore, writing uncivil comments is likely to be seen as a reflection of more stable characteristics of the sender and not as the result of situation-specific aspects that are only temporary.

In this respect, Reeder et al. (2002) demonstrated that people attribute more negative motives and traits to individuals who behave instrumentally aggressive compared with individuals who react to provocations. As incivility is associated with hostile cognitions (Rösner et al., 2016) and negative affect (Chen and Lu, 2017; Gervais, 2017), people presumably attribute aggressive motives to senders of uncivil messages, and nonaggressive motives to senders of messages not containing incivility. Therefore, we expect that ( $H_1$ ) *compared with civil user comments, uncivil comments will (a) positively impact the attribution of aggressive discussion motives and (b) negatively impact the attribution of nonaggressive discussion motives.*

Consequently, attributions of motives should also determine which traits are inferred about senders of uncivil comments (Reeder, 2009). Whereas the attributions of motives depend on people's context-dependent behavior, traits are more stable inferences, helping individuals to predict others' future behavioral patterns (Fiske, 1980). Thus, since individuals attribute more aggressive motives to others who comment in an uncivil way, they may have increased hostile thoughts regarding the commenter, leading to more hostile trait inferences. Accordingly, we assume that ( $H_2$ ) *compared with civil comments, uncivil comments will positively impact the extent of hostile trait inferences through the indirect pathway of attributed motives.*

### ***The role of a comment's stance***

Although incivility within a comment assumably guides which motives and traits recipients attribute to its sender, the literature provides evidence that people process comments differently based on the stance represented therein. Chen and Lu (2017) found that comments arouse negative emotions and aggressive intentions if they directly disagree with the recipient's viewpoints. Likewise, Gervais (2015, 2017) found that compared with like-minded incivility, uncivil disagreement induces feelings of anger and increases aversion toward the sender.

The *self-categorization theory* (Turner et al., 1987), postulates that people favor individuals from social groups to which they feel to belong (i.e. in-groups) and discriminate against individuals they regard as dissimilar (i.e. out-groups). Perceptions of boundaries

between specific social groups and their members can depend on the specific situation and on which social identity is salient. In this vein, attitude (dis)similarity can trigger intergroup differentiation, which can in turn bias attribution processes insofar as people make more unfavorable attributions to individuals with an opposing stance (Kenworthy and Miller, 2002). Moreover, it was found that people attribute more negative motives and dispositions to individuals with dissimilar attitudes than to like-minded individuals (Reeder et al., 2005). When others challenge their views, people can have increased feelings of cognitive dissonance (Festinger, 1957). In such cases, individuals seem to misattribute the resulting unpleasant feelings to the presence of individuals from an out-group, and cope with dissonance by derogating the “others” (Cooper and Mackie, 1983). Consequently, we assume that  $(H_3)$  *compared with like-minded comments, comments representing an opposing stance will positively influence the attribution of aggressive motives for commenting and (b) negatively influence the attribution of nonaggressive motives.*

In line with our aforementioned assumptions, we expect that  $(H_4)$  *compared with comments which support the recipient's stance, comments which oppose the recipient's stance will positively impact the extent of hostile trait inferences through the indirect pathway of attributed motives.*

It remains inconclusive whether effects of incivility are dependent on the sender's stance. Chen and Lu (2017) indicated that both civil and uncivil disagreement increased negative affect and aggressive intentions. In contrast, Gervais (2017) found that incivility only increases feelings of anger when advocating an opposing stance. Likewise, Hutchens et al. (2019) found an interaction effect of group identity and incivility. For feelings of enthusiasm, it did not matter whether comments from in-group members were uncivil or civil. However, compared with uncivil comments, participants reported higher enthusiasm when exposed to civil comments from out-group members. Gervais (2017) argued that, compared with offensive like-minded incivility, like-minded incivility that highlights emotionality might be accepted more by individuals than uncivil disagreement. However, since there is no stable empirical foundation upon which to derive how the stance affects the processing of uncivil comments, we ask (RQ<sub>1</sub>): *Is the attribution process proposed in  $H_1$  and  $H_2$  affected by the comment sender's stance?*

### ***The impact of attributions on discussion behavior***

As argued by the general aggression model (Anderson and Bushman, 2002) people's appraisal of a situation results in more or less thoughtful actions. When enough mental resources are available and the behavior is regarded as important, people are more likely to act thoughtfully and the probability of responding aggressively decreases. Regarding online discussions, it was indeed demonstrated that individuals are less satisfied with the discussion and are less willing to participate therein when it is uncivil (Han and Brazeal, 2015). In this respect, individuals are generally less willing to write comments when they perceive low discussion standards (Springer et al., 2015). Accordingly, we expect that hostile trait inferences will decrease recipients' willingness to join a discussion, because people do not expect an exchange with hostilely assessed individuals to be constructive. Therefore, we assume that  $(H_5)$  *hostile trait inferences will negatively predict recipients' willingness to participate in a discussion.*

Moreover, uncivil statements are more likely to result in uncivil responses than civil statements (Chen and Lu, 2017; Gervais, 2017; Ziegele et al., 2018) which is driven by individuals' experienced anger (Gervais, 2017). We assume that this is the result of hostile trait inferences and therefore hypothesize that ( $H_6$ ) *hostile trait inferences will positively predict recipients' likelihood of uncivil commenting.*

Gervais (2015) argued that criticizing uncivil commenters is a manifestation of aversion, as only unacceptable behavior is criticized. Although such critique can also be expressed in an uncivil manner, it does not necessarily have to be uncivil. When expressed in a civil manner, people might reappraise the situation and, as a result, act in more thoughtful ways (Anderson and Bushman, 2002). Consequently, we also expect that ( $H_7$ ) *the extent of hostile traits that is inferred about the sender of a comment will positively predict recipients' likelihood of expressing critique.*

### **Impact of "dark" personality traits**

So far, we have focused on comment characteristics to explain how people assess its sender. However, the perceiver's personality is important, too. Stable personal characteristics reflect which knowledge structures individuals will consistently use (Anderson and Bushman, 2002). For instance, the dark triad personality traits, a personality model including narcissism (i.e. exaggerated feelings of grandiosity), psychopathy (i.e. lack of affect, self-control, and empathy) and Machiavellianism (i.e. increased tendency to manipulate others for personal gain), have often been found to correlate with various online antisocial behaviors such as cyber-aggression (Moor and Anderson, 2019). Yet, it is largely unclear how these traits influence the appraisal process of antisocial behavior. People high in psychopathy and Machiavellianism have repeatedly been found to lack social cognitive abilities such as empathy (Ali and Chamorro-Premuzic, 2010). Moreover, individuals high in psychopathy were found to have deficits in identifying interpersonal threats (Brankley and Rule, 2014). Notably, the latter study did not find that Machiavellianism affected the perception of threats. However, since Machiavellianism and psychopathy are highly intercorrelated (Ali and Chamorro-Premuzic, 2010), we assume that ( $H_8$ ) *people scoring high both on (a) psychopathy and (b) Machiavellianism will be less likely to attribute aggressive motives for sending uncivil comments compared with people with low levels of these traits.*

In contrast, narcissists are especially vulnerable to negative feedback (i.e. ego-threatening situations), which can result in extreme negative emotions (Li et al., 2016). Therefore, it is likely that individuals high in narcissism attribute more aggressive motives to others whose comments oppose their own stance, because this challenges their ego. Accordingly, it is expected that ( $H_9$ ) *people high in narcissism will be more likely to attribute aggressive motives for sending comments with an opposing stance compared with people with low levels of narcissism.*

### **Method**

Within an online experiment, participants were exposed to a comment on Twitter that either did or did not include incivility and either opposed or corresponded to participants'

own stance. The survey was conducted from 16 August to 16 October 2019 and addressed a German-speaking population. All analyses were conducted using IBM SPSS 26 and IBM SPSS AMOS 26.

### Sample

Participants were mostly recruited on Facebook via groups related to political discussions and via surveycircle.com ( $n=215$ ), an online platform for scientific surveys. Gift cards from an online retail store were offered as an incentive to take part in the study. In total, 469 individuals completed the survey, of which 17 participants were excluded based on unrealistic completion times and/or due to suspicious rating behavior. The final sample therefore included 452 individuals (254 females, one unspecified, four non-binary) aged between 14 and 78 years ( $M=33.10$ ,  $SD=11.89$ ). The majority stated that they had a university degree (61%) and most individuals were students (43%) or employees (39%). Participants located themselves rather on the left-leaning political spectrum:  $M=35.10$  (from 0=*far left* to 100=*far right*). Thirty-four percent of participants indicated that they commented on news on social media or posts of other users at least once a week.

### Procedure and stimulus material

After a briefing, participants indicated their general position toward different social issues on a 7-point semantic differential (*con—pro*). This enabled participants to be allocated to like-minded or opposing comments. Participants who indicated being neither pro nor con regarding the relevant issue were not further surveyed as they could not be assigned to a condition relating to a comment's stance. Remaining participants were randomly assigned to the experimental conditions and were exposed to a mock-up Tweet including one user response. The Tweet was sent from a mock-up Twitter profile which asked recipients for their opinion about a socially relevant issue. The user profile was labeled with a German unisex name and as profile picture a sunset was chosen. Since incivility can have differential effects according to the topic being discussed (Wang and Silva, 2017), we used two different issues to enhance the generalizability of the results. One topic addressed the legalization of marijuana ( $n=227$ ), while the other addressed the introduction of compulsory military service ( $n=225$ ). The accompanying user comment either supported or opposed the issue. Moreover, since Gervais (2017) argued that different forms of incivility (i.e. offending or emotional incivility) can trigger different reactions, the uncivil comment included either name-calling as an offending form of incivility or vulgarity as an emotional form of incivility, as these forms are the most common (Coe et al., 2014). However, the two forms of incivility were merged into a single factor as they were perceived similarly, resulting in a  $2 \times 2$  between-subjects design for the main analyses. The control group was not confronted with any incivility. Depending on the experimental condition, participants were allocated to an uncivil or a civil comment that corresponded with the participant's stance or opposed it (see Table 1 for details).

After participants indicated whether they would participate in the discussion, they were requested to write a comment of their own. Then, in random order, participants

**Table 1.** User comments employed in the study.

	Pro issue (n = 226)	Con issue (n = 226)
No incivility (n = 150)	I'm totally in favor. I think the ban on marijuana is completely unnecessary. But of course it is completely ok if someone else thinks differently	I'm completely against it. I still think the ban on marijuana is seriously necessary. But of course it is completely ok if someone else thinks differently
Name-calling (n = 152) and vulgarity (n = 150)	I'm completely in favor. I think the ban on marijuana is completely unnecessary. At the most, some backward idiots think that we still need it. But these assholes should pipe down	I'm completely against it, damned shit . . . I still think the ban on marijuana is seriously necessary. It makes me want to puke. If someone wants to say the fucking opposite, they should pipe down

Comments are only presented for the topic of marijuana. Name-calling is only presented for the pro issue condition and vulgarity is only presented for the con issue condition.

assessed the manipulated comment according to the sender's potential motives for writing the comment, the sender's traits, and, as a control, the perceived incivility of the comment. Subsequently, participants rated the dark triad personality traits, and provided information about their social media use behavior and sociodemographic characteristics (gender, age, profession, and level of education). In addition, participants rated the relevance of the topic, the sender's perceived intention when writing the comment, and the acceptability of incivility. However, since these variables were not relevant for our hypotheses, they were omitted from the present analyses.

### Measures

When conducting confirmatory factor analyses (CFAs), items with factor loadings smaller than .50 and/or residual covariances greater than .40 were removed (Kline, 2016). The model fit was assessed by common indices from the literature (Byrne, 2010; Hu and Bentler, 1999). Reliability was measured using Cronbach's  $\alpha$ .

**Attributed motives.** Since no measurements exist that operationalize attributed discussion motives, we constructed our own scale based on an explorative approach. To capture nonaggressive motives, we draw on items which were originally developed by Springer et al. (2015) to measure motives related to comment writing. However, the original items were designed for self-assessment. Therefore, we only selected items that we regarded to be transferable to third-person perception and modified them accordingly. Moreover, we only chose items that either focus the goal to discursively exchange with others (seven items from the interactivity dimensions) or on the goal to generate knowledge from the discussion (six items from the cognitive motives dimension) because these motives correspond to an interest in a constructive discussion with others.

To capture aggressive motives, we relied on the *motivation aggression questionnaire* (Ohlsson and Ireland, 2011). However, as the scale was neither developed for third-person perception nor for the online discussion context, items were selectively chosen and

were modified, too. Again, we focused on items we considered to be transferable to others. As the distinction between proactive aggression and reactive aggression is emphasized in the literature (Ohlsson and Ireland, 2011), we constructed items that either addressed motives corresponding to aggression that is instrumental and planned (eight items) or addressed motives corresponding to aggression that is triggered by preceding provocation or threats (seven items).

Importantly, although we have built on existing measurements in the selection of the items, it was a self-constructed scale. Therefore, we used comprehensive statistical analyses to reliably identify distinct factors. An exploratory factor analysis (EFA) with principal component analysis and varimax rotation indicated a four-factor solution, which was confirmed by Horn's (1965) parallel analysis. Based on EFAs with principal axis analysis and promax rotation for a fixed number of four factors, four items with low loadings on the main factor ( $< .50$ ) and/or high loadings on other factors ( $> .20$ ) were removed progressively. Based on the CFA, two items with small factor loadings and 13 items with high standardized residual covariances were removed. This led to a three-factor solution with a good model fit ( $\chi^2(24) = 41.395, p = .015, \chi^2/df = 1.73, CFI = .99, TLI = .99, RMSEA = .04$  (90% CI = [.02, .06]), SRMR = .02). Four items loaded on the factor *Aggression* (e.g. "To disrupt the discussion";  $\alpha = .88$ ), which was further employed as the aggressive dimension of motives. Three items loaded on the factor *Interest* (e.g. "To better understand others" or "because she or he is curious";  $\alpha = .87$ ) and two items loaded on the factor *Counteraction* (e.g. "To counteract inappropriate comments"). As the latter factor only includes two items, we employed the Spearman-Brown coefficient as reliability check indicating a good reliability, too ( $\rho = .79$ ). Motives related to interest and counteraction were employed as nonaggressive motives in the analyses because both can be considered as norm-conforming behavior indicating sincere willingness to discuss with others. Surprisingly, no factor emerged that captured motives related to interpersonal exchange (e.g. "to discuss with others"). However, as the comment used did not ask a question or address another comment, participants probably did not consider such motives as pivotal. All items used for the scale construction are presented in Supplemental Appendix 1 (see Table A1).

**Hostile trait inferences.** Hostile trait inferences were operationalized by the derogation subscale of the *hostile automatic thoughts scale* (Snyder, 1997), which measures hostile thoughts about others including different hostile trait assessments. However, the scale is not limited to trait-related thoughts, but also measures other kinds of hostile thoughts. Given that the multiple inference model (Reeder, 2009) exclusively addresses trait inferences, we only employed five items in the analysis that explicitly refer to traits<sup>1</sup> (e.g. "what an idiot!"), providing a good reliability ( $\alpha = .92$ ). Moreover, the scale correlated strongly ( $r = .91, p < .001$ ) with the other five items ( $\alpha = .90$ ).

**Willingness to comment.** Participants' willingness to comment in the discussion was measured with two items, both rated on a 7-point Likert-type scale. One item asked: "How likely would you be to enter into the shown discussion to express your opinion?" (from 1 = *highly unlikely* to 7 = *highly likely*). The other item asked: "If you had the possibility, how likely would you be to write a comment on this post?" (from 1 = *highly*

*unlikely* to 7=*highly likely*). The scale showed a good reliability (Spearman–Brown’s  $\rho = .87$ ).

**Uncivil commenting and critique expression.** Participants were requested to write a comment regardless of their actual intention to join the discussion. To measure incivility and expressions of critique, a content analysis was performed. Incivility was coded based on conceptualizations commonly employed in the incivility literature (Coe et al., 2014; Stryker et al., 2016). To code expressions of critiques, we relied on Gervais’ (2015: 18) conceptualization of “posts in which specific personal qualities, behavior, and traits of the [comment writer] were negatively assessed.” A second rater, who was unaware of the hypotheses, coded 25% ( $n = 113$ ) of participants’ answers, leading to an intercoder agreement of 98% (Krippendorff’s  $\alpha = .88$ ) regarding uncivil comments and 95% (Krippendorff’s  $\alpha = .80$ ) regarding expressions of critiques. Overall, 27 responses were coded as uncivil and 67 were coded as critique expression.

**Dark triad.** The dark triad traits were assessed based on the German adaptation of the *short dark triad personality measure* (Malesza et al., 2017), which measures each trait with nine items. Items were rated on a 7-point Likert-type scale (from 1=*I do not agree at all* to 7=*I totally agree*). Example items are “Most people can be manipulated” (for Machiavellianism;  $\alpha = .80$ ), “People see me as a natural leader” (for narcissism;  $\alpha = .71$ ), and “People often say I’m out of control” (for psychopathy;  $\alpha = .72$ ).

**Perceived incivility.** As a control, participants rated the comment regarding 21 adjective pairs on a 7-point semantic differential. The adjectives were adapted from the incivility literature (Chen and Ng, 2016; Kenski et al., 2017), but we also added items asking about how stimulating the comment is perceived to be (e.g. “motivating—demotivating”) and the perceived discussion value of the user comment (“constructive—destructive”) to obtain a comprehensive control measure. An EFA with principal component analysis and varimax rotation indicated a two-factor solution. However, we relied on Horn’s (1965) parallel analysis, which indicated a one-factor solution. Based on the CFA, one item with a small factor loading and 13 items with high standardized residual covariances were removed. The model with the remaining seven items provided a good model fit ( $\chi^2(62) = 35.701$ ,  $p = .001$ ,  $\chi^2/df = 2.62$ , CFI = 1.00 TLI = .99, RMSEA = .06 (90% CI = [.04, .08]), SRMR = .01). The final scale included the adjective pairs: “appropriate—inappropriate,” “polite—impolite,” “nice—mean,” “hostile—friendly,” “respectful—disrespectful,” “objective—subjective,” “cooperative—uncooperative.” The scale showed a very good reliability ( $\alpha = .97$ ).

## Results

### *Preliminary analysis*

To check whether people perceived the manipulation in the desired direction and to assess whether there are differences in the perception of name-calling and vulgarity, we conducted an ANOVA including the between-subject variables type of incivility and

sender's stance as independent variables and the perceived incivility as dependent variable. A significant effect of type of incivility on the perceived incivility was revealed ( $F(2, 446) = 569.66, p < .001, \eta^2 = .719$ ). Bonferroni-corrected post hoc tests indicated that participants perceived comments including name-calling ( $M = 44.45, SD = .55$ ) and vulgarity ( $M = 42.14, SD = .55$ ) as significantly more uncivil than civil comments ( $M = 21.16, SD = .55$ , both  $p < .001$ ). Although the difference between name-calling and vulgarity was also significant ( $p = .003, d = 4.2$ ), the effect was relatively small. Moreover, there was no interaction effect between incivility and comment stance ( $F(2, 446) = .99, p = .372, \eta^2 = .004$ ), implying that participants did not perceive vulgarity and name-calling differently based on the sender's stance.

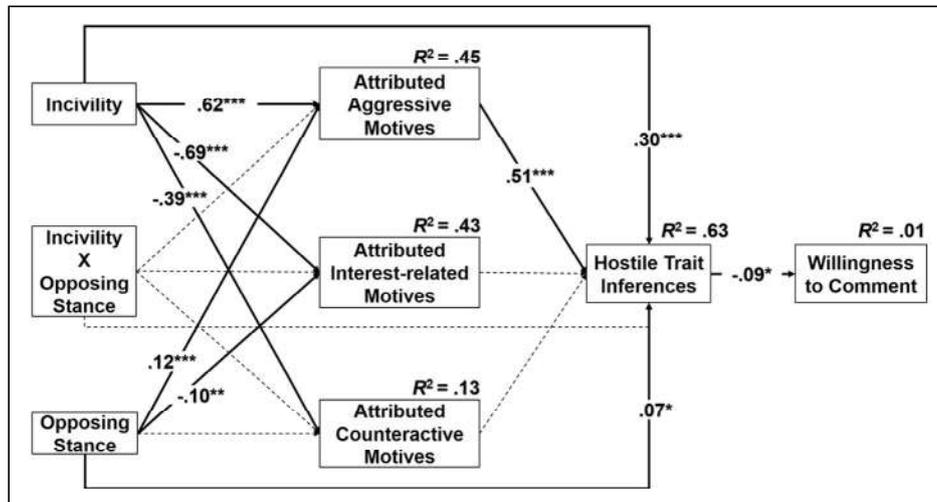
### Attribution process

To test  $H_1$ — $H_5$  and our research question (RQ<sub>1</sub>), a path analysis with manifest variables and maximum likelihood estimation was conducted. The indirect effects were tested by using bias-corrected bootstrapping with 5000 resamples (95% confidence interval). To determine the model fit, common indices were employed. Since extent of incivility and expressions of critique in participants' response comments were coded dichotomously, we additionally computed binary logistic regression models.

As preliminary analyses revealed only minor differences between the perception of name-calling and vulgarity, and no interaction between sender's stance and the different forms of incivility was expected, comments including name-calling and comments including vulgarity were merged into one factor representing uncivil comments. Relying on contrast coding, cases in which participants were exposed to civil comments were coded as  $-1$  and cases in which comments included name-calling or vulgarity were coded as  $.50$ . The model was well-fitting:  $\chi^2(6) = 11.230, p = .082, \chi^2/df = 1.87, CFI = .996, TLI = .983, RMSEA = .044$  (90% CI = [.00, .08]), SRMR = .02. The complete model is visualized in Figure 1.

We assumed that uncivil comments would positively influence the attribution of aggressive motives and negatively influence the attribution of nonaggressive motives ( $H_{1a/b}$ ). Indeed, comments including incivility positively affected the attribution of aggressive motives ( $\beta = .62, p < .001$ ) and negatively influenced the attribution of interest-centered motives ( $\beta = -.69, p < .001$ ) and counteractive motives ( $\beta = -.39, p < .001$ ). We also hypothesized that uncivil comments would positively impact hostile trait inferences through the pathway of attributed motives ( $H_2$ ). Supporting this assumption, aggressive motives positively affected hostile trait inferences ( $\beta = .51, p < .001$ ), and we found a positive indirect effect of uncivil comments on hostile trait inferences ( $\beta = .36, p < .001, CI = [.29, .44]$ ). Notably, neither the attribution of interest-centered motives ( $\beta = -.03, p = .415$ ) nor the attribution of counteractive motives ( $\beta = -.06, p = .076$ ) affected hostile trait inferences, indicating that the indirect effect is mediated by aggressive motives. Moreover, uncivil comments still had a positive influence on hostile trait inferences ( $\beta = .30, p < .001$ ).

Furthermore, we expected that comments representing an opposing stance would positively affect the attribution of aggressive motives and negatively affect the attribution of nonaggressive motives ( $H_{3a/b}$ ). Indeed, we found that comments with an opposing



**Figure 1.** Path model ( $N=452$ ). Significant effects with standardized effect sizes are reported. Model fit indices:  $\chi^2(6) = 11.230$ ,  $p = .082$ ,  $\chi^2/df = 1.87$ , CFI = .996, TLI = .983, RMSEA = .044, SRMR = .02. Coding of incivility: vulgarity = .50, name-calling = .50, and no incivility = -1. Coding of stance: opposing stance = 1 and like-minded stance = 0.  $*p < .05$ ;  $**p < .01$ ;  $***p < .001$ .

stance (compared with a like-minded stance) positively affected the attribution of aggressive motives ( $\beta = .12$ ,  $p < .001$ ) and negatively predicted the attribution of interest-centered motives ( $\beta = -.10$ ,  $p = .005$ ). Stance did not affect participants' likelihood of attributing counteractive motives ( $\beta = -.04$ ,  $p = .405$ ).

Again, we assumed that the comment's represented stance influences the extent of hostile trait inferences indirectly through the attributed motives ( $H_4$ ). The hypothesis was supported, as comments with an opposing stance (compared with a like-minded stance) had an indirect negative influence on hostile trait inferences ( $\beta = .07$ ,  $p < .001$ , CI = [.03, .11]). We also found a direct effect ( $\beta = .07$ ,  $p = .012$ ).

We additionally investigated whether there is an interaction effect between the comment's stance and incivility ( $RQ_1$ ). However, the interaction of the experimental conditions neither affected the motives attributed to the sender (aggressive motives:  $\beta = .05$ ,  $p = .286$ , interest-centered motives:  $\beta = .07$ ,  $p = .161$ , and counteractive motives:  $\beta = .05$ ,  $p = .440$ ) nor the making of hostile trait inferences ( $\beta = .01$ ,  $p = .856$ ). Overall, 45% of the variance in aggressive motives, 43% of the variance in interest-centered motives, 13% of the variance in counteractive motives, and 63% of the variance in hostile trait inferences was explained.

$H_5$  hypothesized that the result of the attribution process would affect people's willingness to join the discussion. As expected, hostile trait inferences negatively predicted participants' willingness to leave a comment in the discussion ( $\beta = -.09$ ,  $p = .047$ ), explaining 1% of the variance.<sup>2</sup> To test whether hostile trait inferences positively affect the usage of incivility ( $H_6$ ) and critique expressions ( $H_7$ ), for each endogenous variable, a binary logistic regression model was calculated. However, when including uncivil

expressions as dependent variable, the overall model was not significant ( $\chi(1)=2.06$ ,  $p=.151$ , Nagelkerke's  $R^2=.01$ ). Consequently, hostile trait inferences did not increase the likelihood of using uncivil expressions ( $b=.03$ , odds=1.03,  $p=.159$ , CI=[.99, 1.07]). By contrast, when adding critique expressions as dependent variable, the overall model was significant ( $\chi(1)=21.637$ ,  $p<.001$ , Nagelkerke's  $R^2=.08$ ). The data revealed an increased likelihood of expressing critique based on inferred hostile traits ( $b=.06$ , odds=1.06,  $p<.001$ , CI=[1.03, 1.09]). It should be noted that Nagelkerke's  $R^2$  was small, indicating a low explanatory power.

### Moderation hypotheses

To calculate the hypothesized moderation effects, we employed the PROCESS macro (model 1) developed by Hayes (2018). Again, civil comments were coded as  $-1$  and comments that included name-calling or vulgarity were coded as  $.50$ . Mean-centered construction of products was employed.

$H_{8a/b}$  assumed that persons who score high on psychopathy and Machiavellianism are less likely to attribute aggressive motives for sending uncivil comments. Supporting  $H_{8a}$ , when psychopathy was included as a moderator, the overall moderation model was significant ( $F(3,448)=124.15$ ,  $p<.001$ ,  $R^2=.45$ ). Uncivil comments ( $b=6.35$ ,  $t(448)=18.70$ ,  $p<.001$ , CI=[5.68, 7.02]) as well as psychopathy ( $b=.07$ ,  $t(448)=2.26$ ,  $p=.024$ , CI=[.01, .12]) predicted the attribution of aggressive motives. Likewise, the interaction of the variables had a significant effect on the attribution of aggressive motives ( $b=-.13$ ,  $t(448)=-3.26$ ,  $p=.001$ , CI=[-.21, -.05]). The 1% increase in explained variance was statistically significant ( $F(1,448)=10.63$ ,  $p=.001$ ,  $\Delta R^2=.01$ ). Using the Johnson–Neyman technique, the data revealed that comments including incivility affected attributions of aggressive motives in 99.8% of cases ( $p\leq .05$ ). However, when individuals had high scores on psychopathy (scores  $>29.57$ ), the effect was no longer significant.

$H_{8b}$  was not confirmed: Although the overall model was significant ( $F(3,448)=115.75$ ,  $p<.001$ ,  $R^2=.44$ ), only uncivil comments predicted the attribution of aggressive motives ( $b=6.36$ ,  $t(448)=18.42$ ,  $p<.001$ , CI=[5.69, 7.04]), whereas Machiavellianism did not affect the criterion ( $b=.04$ ,  $t(448)=1.39$ ,  $p=.164$ , CI=[-.02, .09]). Likewise, no significant interaction emerged ( $b=-.02$ ,  $t(448)=-.56$ ,  $p=.573$ , CI=[-.09, .05]).

$H_9$  claimed that high levels of narcissism would foster the attribution of aggressive motives when people are exposed to a comment representing an opposing stance. The overall model was significant ( $F(3,448)=4.40$ ,  $p=.005$ ,  $R^2=.03$ ). Comments opposing the participant's stance ( $b=1.68$ ,  $t(448)=2.63$ ,  $p=.009$ , CI=[.42, 2.93]), and narcissism ( $b=.09$ ,  $t(448)=2.37$ ,  $p=.018$ , CI=[.02, .17]), positively predicted the attribution of aggressive motives. However, the interaction of the predictors was not significant ( $b=.07$ ,  $t(448)=.92$ ,  $p=.361$ , CI=[-.08, .22]);  $H_9$  was therefore rejected.

## Discussion

The purpose of the current study was to gain knowledge about how people process uncivil discussion comments as social perceivers and how this process affects recipients'

communication behavior. As expected ( $H_1$ – $H_4$ ), we found that, compared with a civil comment, an uncivil comment substantially increased the attribution of aggressive discussion motives and decreased the attribution of nonaggressive motives (i.e. interest-related motives and counteractive motives). Likewise, an opposing comment significantly increased the attribution of aggressive motives and decreased the attribution of nonaggressive motives (i.e. interest-related motives), though the effects were smaller. Moreover, incivility and an opposing stance significantly increased the extent of hostile trait inferences people make about the comment's sender which was also indirectly affected through the pathway of attributed aggressive motives. Again, these effects were considerably stronger for incivility. In turn, hostile trait inferences had a small but significant effect on participants' willingness to write a comment ( $H_5$ ). Furthermore, individuals were more likely to express critique when making higher hostile trait inferences ( $H_6$ ). Against our expectations, individuals were not more likely to comment uncivilly themselves when making increased hostile trait inferences ( $H_7$ ).

The finding that individuals attribute more aggressive motives to others' uncivil behavior is probably triggered by an aversive internal state. Such comments increase recipients' negative emotions and cognitions (see also Chen and Lu, 2017; Gervais, 2017; Rösner et al., 2016), likely leading to more hostile inferences about the sender. Thus, this study demonstrated that uncivil discussion comments crucially guide how people assess its sender. Noteworthy, the comment's stance influenced inferences about the sender's motives and traits regardless of the level of incivility. A potential explanation could be that, due to cognitive dissonance, dissent within a discussion causes negative feelings, which are misattributed to individuals who disagree with one's own stance (i.e. the out-group) to resolve the arising dissonance (Cooper and Mackie, 1983).

The data also revealed that incivility negatively affects the attribution of nonaggressive motives such as interest-based motives (i.e. interest in others opinion and curiosity) and counteractive motives (i.e. countering inappropriate comments). Likewise, the attribution of interest-based motives decreased when the comment included an opposing stance. Thus, comments that are like-minded and civil are probably more likely to cause a positive internal state (e.g. enthusiasm; Hutchens et al., 2019). Interestingly, only incivility affected the attribution of counteractive motives, while the stance did not. Assumably, people expect that individuals who write civil comments aim to prevent subsequent comments from being written in an uncivil manner. However, neither the attribution of interest-related motives nor the attribution of counteractive motives decreased hostile trait inferences. One potential explanation for this tendency is the *negativity bias* which suggests that people give more weight to unpleasant experiences (e.g. uncivil comments) because negative impressions are quicker to form (Baumeister et al., 2001). Therefore, the attribution of aggressive motives may weigh heavier when making trait inferences than the attribution of nonaggressive motives. However, as we did not test such effects, further research is necessary to understand why only aggressive motives affect hostile trait inferences.

We also explored whether there is an interaction effect between a comment's represented stance and the presence of incivility ( $RQ_1$ ). Although both manipulations significantly influenced the attribution process, the interaction neither affected hostile trait inferences nor the attribution of commenting motives. This contrasts with previous

findings (cf. Gervais, 2017; Hutchens et al., 2019). However, these studies manipulated the partisan identification and not, as in the present study, the identification based on participants' opinion toward a socially relevant issue. Disagreement with a topic alone does not seem to affect which attributions individuals make about the senders of uncivil comments. Consistently, Chen and Lu (2017) found that topic-specific disagreement induces negative emotions and aggressive intentions regardless of whether being civil or uncivil. Thus, future studies should be aware that an opposing partisanship is possibly more likely to alter the effects of incivility than issue-specific disagreement.

We further expected that the outcome of the attribution process (i.e. hostile trait inferences) would affect recipients' own communication behavior ( $H_5$ – $H_7$ ). Indeed, participants were less willing to participate in the discussion when inferring higher levels of hostile traits about the sender of the comment. Probably individuals expect the discussion to be less valuable if the discussion partner was assessed hostilely. As people have the tendency to overestimate personality traits or attitudes as reasons for others' behavior (Ross, 1977), they probably consider that the messages might change during the discussion but not the sender. Although this effect was small, given the range of potential factors that can influence whether individuals are willing to participate in a discussion or not, the effect is still significant. Moreover, this finding supports the idea that the attributional perspective can help to explain effects of incivility on its recipients. However, this approach might be especially helpful when investigating outcome variables that are directly related to the comment such as its persuasiveness. Noteworthy, in the current study, participants were only exposed to a single comment whereas, in reality, individuals are usually confronted with several comments simultaneously. Nevertheless, as a first approximation, it is important to understand the effects of a single comment.

The finding that hostile trait inferences only increase the likelihood of expressing critique and not of responding in an uncivil manner gives reason for optimism. From the perspective of the general aggression model (Anderson and Bushman, 2002), participants may have been more prone to make a reappraisal, resulting in a more thoughtful action. Notably, the proportion of participants who frequently comment in online discussions was relatively high in our sample (34%). As such, participants were presumably more likely to consider commenting to be a relevant behavior.

All in all, the findings demonstrate that in online political discussions people do not appraise comments detached from the sender and that both incivility and the stance of a comment can guide how the sender is assessed. People who write uncivil comments online should be aware that they might be assessed by others in an undesired way, including those who are holding the same opinion. Social media and discussion board providers therefore should educate users about these potential consequences. More concerning however is that, although to a lower degree, people also tend to attribute aggressive motives to senders of an opposing stance, no matter whether it is civil or uncivil. Such a tendency is likely to have negative effects on a culture of discussion that should be conducted at eye level. This is all the more important as we found a small but significant negative effect from hostile trait inferences on individuals' willingness to discuss.

We also assumed that individuals who score high on psychopathy and Machiavellianism would attribute less aggressive motives to senders of uncivil comments ( $H_{8a/b}$ ). Indeed, the higher the participants' extent of psychopathy, the less likely they were to attribute

aggressive motives to the sender of the comment. Due to psychopaths' lack of empathy, they are probably unaware that uncivil commenting can have negative effects on others, and thus do not consider it as aggressive behavior. An alternative interpretation is that people with high levels of psychopathy see themselves as superior and, therefore, see others as generally less threatening (Brankley and Rule, 2014). In the context of online discussions, psychopathy might cause individuals to consider others' opinions as irrelevant, regardless of whether or not they are expressed in an uncivil manner. Thus, individuals high in psychopathy seemingly assess incivility differently than people who scored low on this trait. Considering the subsequent effects found in this study, this tendency could also lead to individuals with high levels of psychopathy being more likely to participate in uncivil discussions.

Contrary to our assumption, Machiavellianism did not affect the relationship between uncivil comments and the attribution of aggressive motives. Since people high in Machiavellianism tend to manipulate others for their own personal gain, they might have an enhanced ability to empathize when confronted with personal threats. However, it is largely unknown in which contexts people high in Machiavellianism accurately infer mental states about others and in which contexts they lack this ability (Ali and Chamorro-Premuzic, 2010). This study demonstrated that in the context of discussion incivility, people with high levels of Machiavellianism do not attribute significantly different motives to senders of uncivil messages than individuals low in Machiavellianism and, therefore, differs from individuals high in psychopathy. Further research is needed to understand the role of Machiavellianism in processing uncivil comments.

Furthermore, the effect of comment's stance on the attribution of aggressive motives was not moderated by narcissism ( $H_9$ ). However, narcissism alone predicted the attribution of aggressive motives. In this regard, narcissists tend to anticipate future threats to the ego and are hypervigilant to ambiguous cues of potential threat (Martinez et al., 2008). Therefore, in online discussions, narcissists might expect that others will disagree with them regarding details of the issue or simply interpret comments as disagreeable when they do not exactly reflect their opinion.

The present study gives rise to some practical and theoretical implications. The results suggest that attribution theories (e.g. the multiple inference model; Reeder, 2009) are a promising supplement to the general aggression model (Anderson and Bushman, 2002), helping to explain how people appraise a social situation. Furthermore, the multiple inference model is not only helpful to understand how people explain others' behavior but can also help to predict how people themselves behave (Reeder, 2009). Thus, the behavioral component should be accounted for within the model.

From a practical standpoint, as uncivil comments have the potential to bias attribution processes, senders of uncivil messages should be educated that uncivil commenting in discussions can lead to undesired impression making. A possible approach could be to employ auto-detection interventions. For instance, when using certain words or phrases, people could get informed about the potential effect of incivility on impression making and the recipients' subsequent behavioral intentions. However, it has to be examined how effective such interventions are. Second, to mitigate potential negative effects of incivility and disagreement, that are the result of attribution processes, recipients of discussion comments should be beware that their impressions of others can be biased.

Especially in the case of disagreement people should be more strongly sensitized for this. This could be realized by general guidelines such as netiquettes or by sporadically appearing advices embedded in discussion feeds. Moreover, if automatic detection identifies uncivil language in an initial comment, subsequent commenters could be advised to reconsider their attributions to diminish potential negative reactions.

### *Limitations and further directions*

The generalizability of the findings is limited as the sample is not representative. Participants selected themselves for the study, resulting in a disproportionately high number of young, female and highly educated individuals. However, on a positive note, compared with previous findings (cf. Springer et al., 2015), we succeeded in recruiting a large number of participants who regularly comment online. Furthermore, we did not examine all possible motives and traits that can be attributed to senders of uncivil messages. Likewise, based on the current approach, other outcome variables like perceivers' open-mindedness or perceived polarization could be examined. Moreover, the validity of the self-constructed measurement of attributed motives was not examined. However, this study provides first insights of which motives are attributed to discussion comments. Based on these explorative findings, future studies should refine the operationalization. Likewise, as the scale that measured perceived incivility showed a very high reliability, future research could further explore which items best capture the construct. Moreover, we only employed name-calling and vulgarity as types of incivility in the stimulus material. However, the umbrella of incivility compasses a range of other antisocial behaviors which probably cause different effects. Noteworthy, since participants wrote a comment regardless of their actual intention to do so, some individuals may have written a comment they would not write in reality. Likewise, although online surveys warrant a high degree of anonymity, individuals may have been more willing to respond in a civil manner than in a naturalistic setting due to social desirability bias.

### **Conclusion**

This study has demonstrated that incivility can guide how its sender is assessed. Thus, incivility should not be considered as a simple characteristic of a discussion but rather as social act. Our results indicate that using incivility can crucially bias perceivers' attribution of motives and traits. This is all the more important as the resulting trait inferences significantly influenced individuals' willingness to join the discussion, albeit this effect was small. The same is true for individuals who hold another view than the recipient, though to a smaller extent. People seemingly still have to internalize that stating an opposing opinion in online discussions is not per se aggressively motivated. The results of the study give, however, also cause for optimism: People were not more likely to comment uncivilly themselves when making hostile trait inferences. Rather, they tended to criticize the person or the comment. Nevertheless, as these attributions partly explain why people are discouraged to discuss with others, it should be avoided that people come to such inferences. True to the motto: "Don't be an idiot."

### Acknowledgements

The authors would like to thank the anonymous reviewers for their constructive and helpful feedback.

### Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Note on submission

This manuscript describes original work that has not been published previously and is not under consideration by any other journal. All authors have agreed to the submission.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded by the Digital Society research program funded by the Ministry of Culture and Science of the German State of North Rhine–Westphalia.

### ORCID iD

Jan P Kluck  <https://orcid.org/0000-0001-8432-8705>

### Supplemental material

Supplemental material for this article is available online.

### Notes

1. Items 12, 13, 15, 16, and 18 (Snyder, 1997).
2. Additional calculated path models revealed that the experimental conditions did not directly affect individuals' willingness to comment and, therefore, did not appreciably increase the explained variance (2%). However, when including age, average commenting frequency on social media per week, perceived relevance toward the topic, and attitude toward the topic, variance increased up to 21%, but did not alter the effect of hostile trait inferences on willingness to comment. Thus, the control variables were not included in the final model.

### References

- Ali F and Chamorro-Premuzic T (2010) Investigating theory of mind deficits in nonclinical psychopathy and Machiavellianism. *Personality and Individual Differences* 49(3): 169–174.
- Anderson CA and Bushman BJ (2002) Human aggression. *Annual Review of Psychology* 53(1): 27–51.
- Baumeister RF, Bratslavsky E, Finkenauer C, et al. (2001) Bad is stronger than good. *Review of General Psychology* 5(4): 323–370.
- Brankley AE and Rule NO (2014) Threat perception: how psychopathy and Machiavellianism relate to social perceptions during competition. *Personality and Individual Differences* 71: 103–107.
- Byrne BM (2010) *Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming, Second Edition*. New York: Routledge; Taylor & Francis Group.

- Chen GM and Lu S (2017) Online political discourse: exploring differences in effects of civil and uncivil disagreement in news website comments. *Journal of Broadcasting and Electronic Media* 61(1): 108–125.
- Chen GM and Ng YMM (2016) Third-person perception of online comments: civil ones persuade you more than me. *Computers in Human Behavior* 55: 736–742.
- Chen GM, Muddiman A, Wilner T, et al. (2019) We should not get rid of incivility online. *Social Media + Society*. Epub ahead of print 16 July. DOI: 10.1177/2056305119862641.
- Coe K, Kenski K and Rains SA (2014) Online and uncivil? Patterns and determinants of incivility in newspaper website comments. *Journal of Communication* 64(4): 658–679.
- Cooper J and Mackie D (1983) Cognitive dissonance in an intergroup context. *Journal of Personality and Social Psychology* 44(3): 536–544.
- Festinger L (1957) *A Theory of Cognitive Dissonance*. Palo Alto: Stanford University Press.
- Fiske ST (1980) Attention and weight in person perception: the impact of negative and extreme information. *Journal of Personality and Social Psychology* 38(6): 889–906.
- Gervais BT (2015) Incivility Online: affective and behavioral reactions to uncivil political posts in a web-based experiment. *Journal of Information Technology and Politics* 12(2): 167–185.
- Gervais BT (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.
- Han SH and Brazeal LM (2015) Playing nice: modeling civility in online political discussions. *Communication Research Reports* 32(1): 20–28.
- Hayes A (2018) *Introduction to Mediation, Moderation, and Conditional Process Analysis*. New York: Guilford Press.
- Herbst S (2010) *Rude Democracy: Civility and Incivility in American Politics*. Philadelphia, PA: Temple University Press.
- Horn JL (1965) A rationale and test for the number of factors in factor analysis. *Psychometrika* 30(2): 179–185.
- Hu LT and Bentler PM (1999) Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling* 6(1): 1–55.
- Hutchens MJ, Silva DE, Hmielowski JD, et al. (2019) What's in a username? Civility, group identification, and norms. *Journal of Information Technology & Politics* 16(3): 203–218.
- Hwang H, Kim Y and Kim Y (2018) Influence of discussion incivility on deliberation: an examination of the mediating role of moral indignation. *Communication Research* 45(2): 213–240.
- Jamieson KH and Hardy B (2012) What is civil engaged argument and why does aspiring to it matter? *PS: Political Science & Politics* 45(3): 412–415.
- Kenski K, Coe K and Rains SA (2017) Perceptions of uncivil discourse online: an examination of types and predictors. *Communication Research* 47: 795–814.
- Kenworthy JB and Miller N (2002) Attributional biases about the origins of attitudes: externality, emotionality, and rationality. *Journal of Personality and Social Psychology* 82(5): 693–707.
- Kline RB (2016) *Principles and Practice of Structural Equation Modeling*. 4th ed. New York: The Guilford Press.
- Li C, Sun Y, Ho MY, et al. (2016) State narcissism and aggression: the mediating roles of anger and hostile attributional bias. *Aggressive Behavior* 42(4): 333–345.
- Malesza M, Ostaszewski P, Büchner S, et al. (2017) The adaptation of the short dark triad personality measure—psychometric properties of a German sample. *Current Psychology* 38: 855–864.
- Martinez MA, Zeichner A, Reidy DE, et al. (2008) Narcissism and displaced aggression: effects of positive, negative, and delayed feedback. *Personality and Individual Differences* 44(1): 140–149.

- Moor L and Anderson JR (2019) A systematic literature review of the relationship between dark personality traits and antisocial online behaviours. *Personality and Individual Differences* 144: 40–55.
- Ohlsson IM and Ireland JL (2011) Aggression and offence motivation in prisoners: exploring the components of motivation in an adult male sample. *Aggressive Behavior* 37(3): 278–288.
- Reeder GD (2009) Mindreading: judgments about intentionality and motives in dispositional inference. *Psychological Inquiry* 20(1): 1–18.
- Reeder GD, Kumar S, Hesson-McInnis MS, et al. (2002) Inferences about the morality of an aggressor: the role of perceived motive. *Journal of Personality and Social Psychology* 83(4): 789–803.
- Reeder GD, Pryor JB, Wohl MJA, et al. (2005) On attributing negative motives to others who disagree with our opinions. *Personality and Social Psychology Bulletin* 31(11): 1498–1510.
- Rösner L, Winter S and Krämer NC (2016) Dangerous minds? Effects of uncivil online comments on aggressive cognitions, emotions, and behavior. *Computers in Human Behavior* 58: 461–470.
- Ross L (1977) The intuitive psychologist and his shortcomings: distortions in the attribution process. *Advances in Experimental Social Psychology* 10: 173–220.
- Snyder CR (1997) Assessing hostile automatic thoughts: development and validation of the HAT scale. *Cognitive Therapy and Research* 21(4): 477–492.
- Springer N, Engelmann I and Pfaffinger C (2015) User comments: motives and inhibitors to write and read. *Information, Communication & Society* 18(7): 798–815.
- Stryker R, Conway BA and Danielson JT (2016) What is political incivility? *Communication Monographs* 83(4): 535–556.
- Turner JC, Hogg MA, Oakes PJ, et al. (1987) *Rediscovering the Social Group: A Self-Categorization Theory*. Cambridge, MA: Basil Blackwell.
- Wang MY and Silva DE (2017) A slap or a jab: an experiment on viewing uncivil political discussions on facebook. *Computers in Human Behavior* 81: 73–83.
- Ziegele M, Weber M, Quiring O, et al. (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.

### Author biographies

Jan P Kluck, M. Sc, is a doctoral student at the chair for of Social Psychology: Media and Communication at the University of Duisburg-Essen, Germany. His research explores interpersonal communication with and through digital media. He is particularly interested in the causes and effects of incivility in the political online communication.

Nicole C Krämer is a Full Professor of Social Psychology: Media and Communication at the University of Duisburg-Essen, Germany. She completed her PhD in Psychology at the University of Cologne, Germany, in 2001 and received the Venia Legendi for psychology in 2006. Her research focuses on social psychological aspects of human-machine interaction (especially social effects of robots and virtual agents) and computer-mediated communication (CMC). She investigates processes of information selection, opinion building, and relationship maintenance of people communicating via Internet, especially via social networking sites.

# Article IV

## **Appraising Uncivil Comments in Online Political Discussions:**

### **How do preceding incivility and senders' stance affect the processing of uncivil**

Please note that this is an unreviewed version of the manuscript. The manuscript was revised within a review at Communication Research (conditional acceptance).

Jan P. Kluck\* and Nicole C. Krämer

*Social Psychology: Media and Communication, University of Duisburg-Essen*

### **Acknowledgements:**

We would like to thank Svenja Klocke and Fenja Stratenhoff for their help in collecting data and preparing the stimulus material.

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

### **Declaration of interest:**

None

### **Compliance with Ethical Standard:**

The procedure performed in this study was approved by the ethics committee of the Department of Computer Science and Applied Cognitive Science of the University Duisburg-Essen and is in accordance with the standards of the institutional research committee and with the 1964 Helsinki Declaration, including later amendments or comparable ethical standards.

## **Appraising Uncivil Comments in Online Political Discussions:**

### **How do preceding incivility and the sender's stance affect the processing of an uncivil comment?**

#### **Abstract**

Although the appraisal of online incivility highly depends on the social context in which it occurs, little research has focused on this aspect. Drawing on the General Aggression Model, we assumed that the appraisal of and the reaction to an uncivil discussion comment is affected by its represented stance and the appearance of accompanying comments. To examine these assumptions, we conducted an online experiment ( $N = 611$ ) with a 3 (*uncivil vs. civil vs. no preceding comments*) x 2 (*opposing vs. conforming recipient's views*) between-subjects design. Data revealed that the influence of preceding comments is limited. However, people were more likely to attribute aggressive motives to senders of incivility when it opposed their own opinion. In turn, these attributions increased individuals' anger, anxiety, hostile cognitions, and even enthusiasm. Furthermore, aggressive motive attributions, participants' emotions, and hostile cognitions guided participants' intentions to answer in a discussion-centered and/or aggressive way.

**Keywords:** Incivility, social norms, user comments, online discussions

## Introduction

On social media, citizens can easily engage in political conversations that influence their involvement in online and offline political activities (e.g., Hyun & Kim, 2015). However, discussions on social media are not always constructive. Research demonstrated that social media users often discuss uncivilly (e.g., Su et al., 2018). Incivility has been defined as “features of discussion that convey an unnecessarily disrespectful tone toward the discussion forum, its participants, or its members” (Coe et al., 2014, p. 600). It was found that the exposure to incivility increases individuals’ negative emotions (Chen et al., 2020; Gervais, 2015, 2017) as well as their hostile cognitions (Rösner et al., 2016), polarizes opinions (Anderson et al., 2014), and inhibits open-mindedness towards opposing positions (Borah, 2014; Hwang et al., 2018). Moreover, it was shown that incivility can negatively influence recipients’ behavior as it increases the likelihood of uncivil reactions (Chen & Lu, 2017; Rösner & Krämer, 2016) and discourages individuals from participating in online discussions (Han & Brazeal, 2015).

To explain how incivility affects people in these ways, the *General Aggression Model* (GAM; Anderson & Bushman, 2002) is helpful. The model was developed to explain human aggression comprehensively and postulates that humans develop certain (aggression-related) knowledge structures through their experiences. These structures are connected to individuals’ affective states and guide how people perceive, interpret, and react to their social environment under consideration of situation-related factors. Anderson and Bushman (2002) highlight three knowledge structures that particularly affect the interpretation of aggression in social situations: (1) perceptual schemata (e.g., identifying a comment as an uncivil social act); (2) person schemata (e.g., beliefs about the sender of a comment); and (3) behavioral scripts (e.g., information about how individuals should react to a discussion comment).

Therefore, if people regard a comment written by specific persons as uncivil, their reaction is guided by behavioral scripts they have learned through past experiences (e.g., reacting uncivilly or refraining from commenting). However, research on incivility faces challenges because the dynamics in online discussions can be very complex. For instance, when approved by all communication partners, incivility can even serve positively as a social function that helps to manage relationships (Chen et al., 2019). Thus, the consequences of uncivil user comments depend strongly on the social context. As incivility in most studies was treated as an attribute of a discussion that is either present or not, at this time, little is known about how people appraise specific uncivil messages and how the social context alters this appraisal. This is important because, in online discussions, individuals are typically confronted with several comments of very different types. Whereas some comments might be uncivil, others are probably written in a civil manner. Especially in public online discussions, people are usually exposed to heterogeneous opinions. Thus, to make an impression of a discussion, people must process various comments that can be very distinct. In this respect, it has not been confirmed that the negative effects of incivility linearly increase with the amount of incivility. Quite the contrary, when a discussion is solely uncivil, negative consequences of incivility might even drop (Rösner et al., 2016). Therefore, it is important to examine how people assess single uncivil comments in the social context to disentangle how incivility has undesired outcomes.

To explain how individuals assess and react to uncivil messages based on contextual factors such as the extent of incivility in accompanied comments or the stance of a comment, the GAM (Anderson & Bushman, 2002) provides a valuable theoretical framework. We conducted an online experiment expecting that the extent of incivility in preceding comments and the opinion of the sender alter how an uncivil comment influences recipients' emotions and cognitions. In turn, we expected that recipients' emotions and cognitions guide the

interpretation of senders' commenting motives which lead to more or less aggressive reactions. The study has been preregistered within the open science framework (OSF<sup>2</sup>).

### **Incivility and the General Aggression Model**

The GAM (Anderson & Bushman, 2002) has already been applied to different contexts, such as physical violence or the consumption of violent media (Allen et al., 2018). A key component of the GAM is the assumption that people develop certain knowledge structures through their experiences which guide how people assess and react to aggressive events. These knowledge structures can be activated more or less automatically and are triggered by certain stimuli (e.g., violent media) or personal characteristics (e.g., beliefs and attitudes). Anderson and Bushman (2002) named three knowledge structures as particularly important: (1) Perceptual schemata guide how individuals categorize certain events. In the context of online discussions, this can be the identification of a discussion comment as a comment but also to identify a comment as uncivil. (2) Person schemata represent the beliefs about others, including social groups. For instance, in online discussions, that can be beliefs about persons with a specific opinion or affiliation to a certain party. (3) Behavioral scripts guide people's behavior under given circumstances. Thus, behavioral scripts can determine in which way individuals react to other (uncivil) comments. Each of these three structures is elementary as they guide how situational aspects influence individuals' internal state, which is characterized by their affect, arousal, and cognitions. Indeed, there is broad evidence that incivility influence individuals' internal state variables.

Affect represents individuals' emotions and moods. Whereas the GAM (Anderson & Bushman, 2002) focuses on negative emotions like anger or hostility, in the political context, three distinct emotional states are highlighted in the literature (Marcus et al., 2000, 2006): enthusiasm, anxiety, and aversion. Enthusiasm reflects individuals' willingness to like and

---

<sup>2</sup> [https://osf.io/m67u8/?view\\_only=e6691f4455ad478ca6ab4bf69923e354](https://osf.io/m67u8/?view_only=e6691f4455ad478ca6ab4bf69923e354)

engage with political stimuli, whereas aversion relates to individuals' intention to avoid hated stimuli (Marcus et al., 2006). Anxiety is particularly important in novel situations as it helps individuals to deal with surprising and threatening stimuli (Marcus et al., 2000). Indeed, all three emotions have already been found to be affected by incivility. For instance, it was found that people feel higher levels of enthusiasm when exposed to civil discussion comments compared to uncivil discussion comments (Hutchens et al., 2019). Likewise, it was shown that feelings of aversion in the form of anger increase when individuals are exposed to incivility (Chen et al., 2020; Gervais, 2015, 2017). However, different findings were made regarding anxiety. Whereas Gervais (2017) found that exposure to incivility reduces feelings of anxiety, Lu & Myrick (2016) found that disagreeable uncivil comments increase recipients' feelings of anxiety. Since anxiety, in the first place, increases to heighten attention and consciousness in new and threatening situations (Marcus et al., 2000), it probably depends on the context of whether incivility positively or negatively affects anxiety.

Apart from individuals' affect, their cognitions are a central part of the internal state. Within the GAM, hostile cognitions are considered as the "acceptability of aggressive concepts in memory" (Anderson & Bushman, 2002, p. 38). In this respect, Bushman and Anderson (2002) demonstrated that people who played a violent video game are more likely to indicate aggressive thoughts within a story completion task. Likewise, Rösner and colleagues (2016) found that exposure to uncivil discussion comments increases recipients' hostile cognitions.

In the next stage, the model suggests that individuals' emotions and cognitions determine how people appraise a social (aggressive) situation which in turn guides their behavior. How people appraise a social situation can be explained by attribution processes that usually take place when people try to make sense out of others' behavior. Anderson and Bushman (2002) propose that people often make certain trait or situation inferences when

explaining others' behavior. However, attribution theorists highlight that this person–situation dichotomy is only useful to explain accidental behavior or outcomes of actions (e.g., success or failure) because in these cases, actors lack direct control. Intentional behavior, in contrast, is explained by the intentions and traits that are inferred to an actor (Jones & Davis, 1965; Malle, 2011; Reeder, 2009). Especially, the motives people attribute to other persons seem to be important. To interpret another persons' intentional behavior, people usually search for the actor's motives to understand the reasons behind their behavior. By doing so, individuals also consider the influence of certain situation-related aspects, such as the behavior of bystanders and other actors (Reeder, 2009; Reeder & Trafimow, 2004). Accordingly, we argue that the motives that are attributed to the sender of an uncivil message play an important role in assessing others' discussion behavior.

It is important to note that although the GAM (Anderson & Bushman, 2002) proposes that processes related to the internal state variables (i.e., affect, arousal, and cognitions) precede the appraisal process. It also acknowledges that a reappraisal can change an individuals' initial rudimentary internal state. A reappraisal is likely if people are motivated to appraise a situation accurately and have enough resources to do so. This is an important aspect as other scholars even argue that making attributions is a precondition for feeling emotions like anger (e.g., Clore & Centerbar, 2004). Thus, how people appraise a situation, driven by attribution processes, is guided by individuals' initial emotions and cognitions. However, in turn, the appraisal can alter individuals' initial emotions and cognitions when people are motivated to think about a situation (see also Berkowitz, 1990).

Finally, the appraisal of a situation determines how people react. Depending on individuals' mental resources and motivation, this can be thoughtful or impulsive and, in both cases, depending on the behavioral scripts that are activated (aggressive or unaggressive; Anderson & Bushman, 2002). For online discussions, it has been repeatedly demonstrated

that people are more prone to react uncivilly themselves when exposed to uncivil comments compared to more civil ones (e.g., Chen & Lu, 2017; Rösner & Krämer, 2016).

### **Impact of Preceding Comments**

The GAM (Anderson & Bushman, 2002) provides an integrative and comprehensive framework that helps to explain how incivility affects individuals and how the context can alter this process. Therefore, we assume that the model provides insights into how the effects of uncivil discussion comments can differ based on which perceptual schemata are activated (i.e., how a social event is categorized). We expect that the process proposed by the GAM alters depending on whether comments that precede an uncivil comment are also uncivil or not. Rösner and colleagues (2016) found that individuals' hostile cognitions increase when being exposed to uncivil comments compared to civil comments. However, participants' hostile cognitions did not increase linearly to the proportion of uncivil comments. Quite the contrary, comparing hostile cognitions when exposed to one, three, or six uncivil comments (out of six discussion comments), no significant difference was found. Data indicated that the impact of incivility on an individual's cognitions even dropped when all comments were uncivil. Thus, more incivility does not automatically lead to more negative internal states. This indicates that the proportion of incivility in a discussion might determine whether incivility is perceived as a norm violation or not.

This pattern could be explained by the *Focus Theory of Normative Conduct* (Cialdini et al., 1990). This theory distinguishes descriptive norms and injunctive norms. Descriptive norms refer to the perception of what is typically done and drive individuals' behavior "by providing evidence as to what will likely be effective and adaptive action" (p. 1015). Injunctive norms refer to assumptions about which behavior is approved or disapproved by others. These norm perceptions can co-exist, but which norm has more impact on individuals' behavior crucially depends on which norm is more salient. In the context of uncivil

comments, people probably consider such behavior as antinormative because they follow an injunctive norm (i.e., incivility is societally disapproved). However, if all comments in a discussion are uncivil, the descriptive norm might mitigate the influence of the injunctive norm because, in this context, incivility might be regarded as normative behavior. Therefore, the effects of incivility on anger (Chen et al., 2020; Gervais, 2015, 2017) and hostile cognitions (Rösner et al., 2016) should be weaker when individuals are exposed to an uncivil comment that is preceded by other uncivil comments. Likewise, the negative effect of incivility on enthusiasm (Hutchens et al., 2019) should decrease in that context. Moreover, levels of anxiety should decrease when individuals are exposed to an uncivil comment that is preceded by other uncivil comments because an uncivil comment that is preceded by civil comments might occur more unexpectedly (see Marcus et al., 2000, 2006). Therefore, it is hypothesized: *H1.1: When exposed to an uncivil comment which is accompanied by preceding uncivil comments, people will have (a) fewer feelings of anger, (b) anxiety, (c) higher levels of enthusiasm, and (d) fewer hostile cognitions compared to an uncivil user comment accompanied by preceding comments not including incivility.*

Although little is known about the motives attributed to senders of uncivil messages, it was already demonstrated that people attribute more negative motives to individuals who behave proactively aggressive compared to individuals who aggressively react to offensive behavior (Reeder et al., 2002). Therefore, we assume further: *H1.2: When exposed to an uncivil comment which is accompanied by preceding uncivil comments, people will attribute less aggressive motives to the comment's sender compared to the sender of an uncivil comment which is accompanied by preceding comments, not including incivility.*

Further, we expect that individuals' internal state has a mediating function between the input variables and the appraisal process. Thus, it is assumed that: *H1.3 Preceding uncivil comments have an indirect negative effect on the attribution of aggressive motives through*

*the pathway of (a) feelings of anger and anxiety, (b) levels of enthusiasm, and (c) hostile cognitions.*

### **Influence of Comment Sender's Opinion**

As proposed by the GAM (Anderson & Bushman, 2002), person schemata seem to be important for how people process incivility, too. It was shown that both civil and uncivil comments arouse negative emotions and aggressive intentions when they contradict the opinion expressed by recipients (Chen & Lu, 2017). Moreover, it was found that individuals have increased feelings of anger and aversion when they are confronted with incivility from an opposing social group (Gervais, 2015, 2017). Likewise, Hutchens and colleagues (2019) found that the negative effect of incivility on enthusiasm can only be found when a comment is written by people from a social group with different attitudes. Thus, the stance of the sender can predict how people react to uncivil messages.

This tendency can be explained by the *Self-Categorization Theory* (Turner et al., 1987). The theory postulates that people assess individuals from like-minded groups more positively than individuals from a group to which they feel not belonging. Noteworthy, feelings of group belongingness and individuals' personal identity can co-occur. Whether people perceive boundaries to others depends on which identity is salient. Therefore, we expect that individuals are more likely to have a negative internal state when confronted with oppositional stances in discussions. Although it was not confirmed that anxiety increases when exposed to civil or uncivil disagreement compared to a neutral control (cf. Gervais, 2017), we still assume that incivility opposing one's view increases feelings of anxiety when directly compared to like-minded incivility, as the latter one should be experienced as less threatening. Therefore, we assume that: *H2.1: When exposed to discussion comments opposing one's stance, people will have higher feelings of (a) anger, (b) anxiety, (c) lower*

*levels of enthusiasm, and (d) more hostile cognitions compared to discussion comments with a like-minded stance.*

Moreover, it was shown that people attribute more negative motives to individuals with dissimilar attitudes than to like-minded individuals (Reeder et al., 2005). Being exposed to opposing opinions challenges one's own attitudes. People seem to misattribute the resulting negative feelings to the "other side" to restore cognitive consistency by derogating the respective out-group (Cooper & Mackie, 1983). Thus it is expected that: *H2.2: When exposed to an uncivil comment opposing one's stance, people will attribute more aggressive motives to the sender of the comment compared to an uncivil comment with a like-minded stance.*

Again, if people's internal state is rather negative when exposed to opposing views, these cognitions and feelings should also affect the motives for sending an uncivil comment. Therefore, we expect that: *H2.3 The exposure to comments with an opposing stance has an indirect positive effect on the attribution of aggressive intentions through the pathway of (a) feeling of anger, (b) anxiety, (c) levels of enthusiasm, and (d) hostile cognitions.*

### **Interaction of Preceding Comments and Stance of the Comment**

The perception of social norms is affected by processes of self-categorization because in-group members have a stronger prescriptive value than out-group members (Hogg & Reid, 2006). This suggests that people are more willing to accept incivility as a norm when they are exposed to a comment from an in-group member, which is preceded by like-minded incivility and not by opposing incivility. Consequently, they should feel less aversion, more enthusiasm, and less hostile cognitions. Otherwise, individuals do not conform to a perceived social norm from an in-group because they think it is the right thing to do but to comply with empirical and normative expectations (Bicchieri, 2017). Thus, they might be more likely to conform to behavior, but their emotions and hostile cognitions are not affected by an

interaction of the experimental conditions. Since empirical evidence is scarce, we ask: *RQ1: Will there be an interaction between incivility in preceding comments and comments' stance?*

### **Behavioral Intentions of Recipients**

When people have appraised a certain (aggressive) situation, they decide how to react (Anderson & Bushman, 2002). Hence, we expect that individuals activate certain behavioral scripts when perceiving hostility which in turn leads to more aggressive behavior. When people assess a comment to be aggressively motivated, we expect that this appraisal activates aggressive behavior scripts. Therefore, people should have a stronger intent to answer aggressively to the uncivil comment (see also Chen & Lu, 2017) and less the intention to discuss constructively with the person. Thus, we expect that: *H3: The attribution of aggressive intent (a) negatively affects an individual's intentions to write a discussion-centered answer and (b) positively affects an individual's intentions to write an aggressive answer to the respective comment.*

It was repeatedly demonstrated that people are more likely to answer in an uncivil way when confronted with incivility in discussions. For instance, Chen and Lu (2017) found that uncivil disagreement, but not civil disagreement, increased uncivil responses. Moreover, Gervais (2017) found that compared to like-minded incivility, only incivility from an opposing group fostered uncivil answers. This indicates that people are more likely to react uncivilly to incivility when they feel threatened. However, Gervais (2015) postulates that criticizing others is also a good indicator for an anger-induced reaction. Indeed, some people probably have learned that criticizing others in a civil manner is a more appropriate reaction to aggression. Thus, although critique can be uncivil, it does not have to be. Therefore, we argue that: *H4: Attribution of aggressive discussion motives positively affects an individual's likelihood of (a) writing an uncivil answer and of (b) criticizing the sender of the comment.*

Noteworthy, although not hypothesized in the preregistration, from a theoretical perspective, the strength of individuals' behavioral intentions guides whether people act accordingly. Ajzen (1991) argues that "the stronger the intention to engage in a behavior, the more likely should be its performance" (p. 181). As aggressive intentions can be considered as a precursor of the discussion behavior addressed in H4, the relationship is probably mediated by individuals' own aggressive answering intentions. Therefore, it is also explored whether such mediating effects occur. The conceptual model visualizing all hypotheses is presented in *Figure 1*.

**[Insert Figure 1 here]**

### **Method**

To test the hypotheses, an online survey addressing a German-speaking population was conducted. A single uncivil statement has been constructed to serve as a basic uncivil comment (i.e., the comment to be evaluated). According to the experimental conditions, the basic uncivil comment was preceded by either five civil comments, five uncivil comments, or, as a control, by no comments. Moreover, comments were consistently either pro or con the issue. Data analyses were conducted with IBM SPSS 25 and IBM SPSS Amos 25.

### **Sample**

In total, 651 participants completed the survey. Individuals were primarily recruited via social media platforms. However, some individuals have additionally been recruited via surveycircle.com ( $n = 101$ ) and prolific.com ( $n = 49$ ), which are platforms for scientific survey recruitment. As an incentive to take part in the study, gift cards for retail stores were raffled. In total, 40 participants were excluded from the analyses based on unrealistic completion times and incomprehensible answering behavior (e.g., when answering inversed items). Thus, the final sample included 611 individuals (397 females, 6 not specified) who were on average 32 years old (ranging from 18 to 76). Most participants indicated that they had a university degree (47.5%) or graduated from high school (29.8%) and were either

university students (43.5%) or employees (36.8%). Moreover, individuals stated that they use social media on average 2.22 hours ( $SD = 1.76$ ) per day and 32.9% of participants indicated to write a discussion comment on social media at least once a week.

### **Stimulus Construction and Pre-Test**

The comments to which the participants were exposed were embedded in a Facebook post that neutrally asked for others' opinions towards a socially relevant topic. To enhance the generalizability of the findings, we selected two different issues as it was shown that the topic of discussion can alter the effects of incivility (Wang & Silva, 2017). We considered both topics as socially relevant, and both topics are regularly discussed on social media. One issue has addressed the question of whether school canteens should be obliged to offer only vegetarian food ( $n = 307$ ), and the other issue has addressed the question of whether fees for public broadcasting services should be abolished ( $n = 304$ ). To avoid undesired effects of names or profile pictures, these features were blurred. In all experimental conditions, participants were exposed to one uncivil comment that was either preceded by five uncivil comments, five civil comments, or blurred comments as a control. These comments were consistently either pro or con the issue addressed by the Facebook post. Examples of the comments are presented in *Table 1*

**[Insert Table 1 here]**

To ensure that the uncivil version of the preceding comments and the basic uncivil comment is perceived in a similar uncivil manner, we conducted a pre-test ( $N = 30$ ). Therefore, all uncivil comments employed in the study have been evaluated a priori. To have comparative values, three extreme uncivil and three civil comments were additionally included in the pre-test. Results revealed that the six comments intended for the study were not perceived differently uncivil. However, the six comments differed significantly from the

civil comments and the extreme uncivil comments as intended. For the data analysis of the pre-test, please see online Appendix A.

## Measures

*Emotions.* Individuals' enthusiasm, anger, and aversion were each measured by items recommended by Marcus and colleagues (2006). For each emotion, we supplemented some items to receive a more comprehensive measurement. Individuals indicated how much they feel certain states on a 7-point Likert scale (from 1 = *not at all* to 7 = *very*). For enthusiasm, individuals rated on *enthusiastic, hopeful, proud, interested, elated, motivated, activated* and *encouraged* (Cronbach's  $\alpha = .90$ ). For anger, individuals rated on *hatred, contempt, bitterness, resentful, furious, angry, and hostile* ( $\alpha = .92$ ), and for anxiety, individuals rated on *anxious, worried, afraid, unsettled, nervous, frightened* ( $\alpha = .91$ ).

*Hostile cognitions.* Hostile cognitions were measured by the approach employed by Bushman and Anderson (2002). Participants read an ambiguous story stem about a car accident and were asked to indicate open-ended ideas about what the protagonist might do, feel, think, or say in the further course. It must be noted that this task originally includes two further stories. However, as Rösner and colleagues (2016) found that only the story with the car accident (i.e., the first story) captured individuals' hostile cognitions, we only adapted this story for the current study. Individuals could indicate up to 20 ideas on how the story could continue. To measure individuals' hostile cognitions, the percentage of answers including aggressive acts, thoughts, or feelings was calculated from the total number of individuals' answers. Overall, 16.42% of the answers were coded as aggressive. A second rater, blind to the hypotheses and conditions, also coded the responses. Following the previous approaches (Bushman & Anderson, 2002; Rösner et al., 2016), we calculated the intraclass correlation coefficients (ICC) for the sum scores of aggressive answers indicating good reliability (ICC = .86; Cicchetti, 1994).

*Attributed aggressive motives.* As no standardized measurements exist, we captured the attributed aggressive motives with nine self-constructed statements which were rated on a 7-point Likert scale (from 1 = *I do not agree at all* to 7 = *I totally agree*). An exploratory factor analysis (EFA) with a principal component analysis (PCA; Varimax rotation) revealed a two-factor solution that was supported by Horn's (1965) parallel analysis. Based on an EFA with a principal axis analysis (PAA; Promax rotation) for a fixed number of two factors, two items with low loadings on the main factor ( $< 0.50$ ) were removed. The first factor included five items that addressed aggressive attributions. The second factor included two items that addressed positive motives, which should serve as reserved items. As the second factor did seemingly not capture aggressive attributions (e.g., "...to encourage other people") only the first factor (e.g., "...to unsettle other people" or "...to attack other people;"  $\alpha = .80$ ) was employed in the analyses. All items and factor loadings are presented in online Appendix B.

*Intentions to react.* To capture individuals' aggressive and discursive intentions to react to the comment, we constructed eight items each that were rated on a 7-point Likert scale (from 1 = *I do not agree at all* to 7 = *I totally agree*). An EFA with PCA (Varimax rotation) supported a two-factor solution that was also confirmed by Horn's (1965) parallel analysis. Based on a further EFA with PAA (Promax rotation) for a fixed number of two factors, only one item was excluded from the scale because of high loading on the other factor ( $>0.20$ ) leading to the anticipated factors aggressive (e.g., "...to signal the person to shut up" or "...to humiliate the person;"  $\alpha = .93$ ) and discursive intent (e.g., "...to better understand the opinion of the person" or "...to share your knowledge with the person;"  $\alpha = .87$ ). Again, items and factor loadings are presented in online Appendix B.

*Uncivil commenting and critique expression.* To measure the likelihood of uncivil commenting and critique expressions, individuals were asked to write a comment regardless of their actual intention to do so. Answers were coded based on markers as proposed in

incivility literature (Coe et al., 2014; Stryker et al., 2016). Moreover, comments were coded as critique expression when they negatively assessed “specific personal qualities, behavior, and traits” of the commenter (Gervais, 2015, p. 18). 8.67% of the responses were identified as uncivil and 61.54% were identified as critique expression. Again, a second rater, blind to the assumptions, coded 10% of answers, leading to a satisfactory intercoder agreement (uncivil comments: Krippendorff’s  $\alpha = .93$  and critique expressions: Krippendorff’s  $\alpha = .83$ ).

*Control.* To test whether people perceived the basic uncivil comment as well as the preceding comments in the intended direction, individuals assessed both on a 7-point semantic differential including five adjective pairs reflecting perceived incivility (*polite – impolite, respectful – disrespectful, friendly – hostile, adequate – inadequate* and *nice – nasty*). Both scales provided a good reliability (basic uncivil comment:  $\alpha = .88$  and preceding comments:  $\alpha = .98$ ). Individuals in the control condition only rated on the basic uncivil comment.

## **Procedure**

Within the survey, participants first indicated their opinion towards the two discussion topics. In addition to specific aspects of the topics, they also indicated their general opinion on the topics on a semantic 7-point differential (con - pro). Based on the answers, in accordance with the experimental condition, individuals were either assigned to like-minded or opposing comments. Individuals who were neither pro nor con regarding the relevant topic were excluded from the survey. The remaining participants were then allocated to the stimulus according to the experimental condition they were assigned to. Then, they indicated their willingness to join the discussion. Afterward, individuals answered open questions as part of a story completion task and indicated their emotional state. Afterward, individuals were again exposed to the stimulus, but this time they were asked to focus on the last comment (i.e., the basic uncivil comment). In random order, individuals assessed the

comment writer's motives and indicated how they would react to the comment. Thereafter, they rated on the perceived incivility of the evaluated comment and preceding comments. Last, individuals rated their characteristics<sup>3</sup> and indicated some sociodemographic information.

## Results

### Manipulation Check

An independent sample t-test revealed that uncivil preceding comments ( $M = 29.19$ ,  $SD = 6.14$ ) were perceived as more uncivil than the civil preceding comments ( $M = 11.91$ ,  $SD = 6.02$ ;  $t(407) = 28.73$ ,  $p < .001$ ). Moreover, a pair-wise t-tests indicated that the civil preceding comments were perceived as less uncivil ( $M = 11.91$ ,  $SD = 6.02$ ) than the basic uncivil comment ( $M = 31.36$ ,  $SD = 3.90$ ;  $t(199) = 34.65$ ,  $p < .001$ ,  $d = 3.84$ ). Against our intention, uncivil preceding comments were perceived as slightly less uncivil ( $M = 29.19$ ,  $SD = 6.14$ ) than the basic uncivil comment ( $M = 31.70$ ,  $SD = 4.24$ ,  $t(208) = 6.67$ ,  $p < .001$ ,  $d = .48$ ). However, the difference was comparatively small and therefore negligible.

### Hypotheses test

A path analysis was conducted using SPSS Amos 25 to test the theoretical model as proposed in H1.1a – H3b and to answer RQ1. Indirect effects were examined by using bias-corrected bootstrapping with 5000 resamples (95% confidence interval). To determine the model fit, we followed indices proposed in the literature (Byrne, 2010; Hu & Bentler, 1999). As the individuals' response comments were coded dichotomously, the effects of aggressive motive attribution on the likelihood to criticize or to comment uncivilly (H4a/b) and the mediating effects of aggressive answering intent were tested employing the Process macro for SPSS (Hayes, 2018), which automatically uses logistic regressions for binary dependent variables.

---

<sup>3</sup> Please note that we have also generated some moderation hypotheses within the preregistration. However, as the moderation analyses did not provide notable additional explanatory value, the result are solely presented in the online Appendix C

In a first step, we tested the path model as proposed in the hypotheses H1.1a – H3b. The experimental conditions were included as dummy coded variables. As the factor preceding comments has been multicategorical, it was included in the model by employing indicator coding (Hayes & Montoya, 2017). In this way, the uncivil preceding comment condition and the control condition (no preceding comments) were each tested against the civil comment condition to examine whether effects are caused by preceding uncivil comments or based on the absence of civility. For the interaction term, the dummy-coded variables uncivil preceding comments and opposing stance were multiplied. Correlations of the metric variables are presented in *Table 2*.

**[Insert Table 2 here]**

When calculating the path model as proposed, it poorly fitted the data:  $\chi^2 (18) = 140.105, p < .001, \chi^2/df = 7.78, CFI = .91, TLI = .74, RMSEA = .11$  (90% CI = [.09, .12]), SRMR = .07. The modification indices provided by AMOS 25 indicated that adding paths from the variables opposing stance, the interaction term, anger, and hostile cognitions to the dependent variable aggressive answering intentions would improve the fit. We also added respective path from the uncivil preceding comments condition as only then the interaction can be meaningful. However, the model still did not acceptably fit to the data:  $\chi^2 (13) = 80.851, p < .001, \chi^2/df = 6.219, CFI = .95, TLI = .80, RMSEA = .09$  (90% CI = [.07, .11]), SRMR = .05.

Since the data did not fit our hypothesized model, based on theory, we slightly adapted the proposed model: As explained by the GAM (Anderson & Bushman, 2002), persons' internal state and the appraisal of a situation do not follow a clear hierarchy. An initial rudimentary internal state can also be overridden by the subsequent appraisal process when people are motivated to think about a situation (Anderson & Bushman, 2002; Berkowitz, 1990). In this vein, individuals were aware of the study situation and were

explicitly asked to focus on the comments. It is therefore conceivable that individuals were motivated to accurately make sense out of the comments and therefore made a reappraisal which has overwritten rudimentary emotions and cognitions. Consequently, in the alternative model, we have replaced the aggressive motive attributions before the internal state variables in the way that the experimental conditions affect the aggressive motive attributions, which in turn influences the internal state variables. However, both the attributed motives as well as the internal state variables were assumed to affect individuals' intentions to respond as these variables presumably interactively influence persons' behavior. Indeed, the alternative model fitted the data considerably better:  $\chi^2(26) = 52.201, p = .002, \chi^2/df = 5.779, CFI = .95, TLI = .96, RMSEA = .04$  (90% CI = [.06, .82]), SRMR = .04. Therefore, this model was employed for the analyses. The final model is visualized in *Figure 2*.

**[Insert Figure 2 here]**

As the alternative model did not include the paths between the experimental conditions and the internal state variables, H1.1a – d and H2.1a – d could not be examined. Consequently, the indirect effects of internal state variables on the attribution of aggressive motives (H1.3a – d and H2.3a – d) could not be observed. However, we also hypothesized that preceding uncivil comments (H1.2), as well as an opposing stance (H2.2), will positively affect to which extent individuals attribute aggressive motives to the sender of the basic uncivil comment. Whereas, neither uncivil preceding comments ( $\beta = .00, p = .963$ ) nor the control condition ( $\beta = .08, p = .148$ ) affected the extent of aggressive attribution making, H1.2 was rejected. However, as expected (H2.2), individuals attributed more aggressive motives to the basic uncivil comment when they were confronted with opposing stances compared to like-minded stances ( $\beta = .12, p = .014$ ). Nevertheless, the interaction of preceding comments and opposing stances did not influence the extent of aggressive motive attributions (RQ1;  $\beta = -.03, p = .631$ ). Although not hypothesized, we found that attributing

aggressive motives to the sender of an uncivil message increased individuals' anger ( $\beta = .11$ ,  $p = .005$ ), anxiety ( $\beta = .19$ ,  $p < .001$ ), enthusiasm ( $\beta = .19$ ,  $p < .001$ ), and hostile cognitions ( $\beta = .12$ ,  $p = .004$ ). This picture reflects in part the relationships expected by the indirect effects expected in H1.3a – d and H2.3a – d. Surprisingly, however, feelings of enthusiasm also increased when higher levels of aggressive motives were attributed to the sender of the basic uncivil message. Furthermore, as expected we found that attributing aggressive motives increased individuals' intentions to answer aggressively (H3b;  $\beta = .24$ ,  $p < .001$ ) but did not influence intentions to answer in a discussion-centered manner (H3a;  $\beta = -.04$ ,  $p = .291$ ). However, aggressive motive attributions had a positive indirect effect on both answering intentions through the pathway of internal state variables ( $\beta = .07$ ,  $p = .003$ ,  $CI = [.05, .11]$  for discussion-centered intentions and  $\beta = .05$ ,  $p = .009$ ,  $CI = [.03, .09]$  for aggressive intentions).

Moreover, although the alternative model did not include direct paths of the experimental conditions on the internal state variables, we found that an opposing comment stance indirectly increased anger ( $\beta = .01$ ,  $p = .006$ ,  $CI = [.00, .04]$ ), anxiety ( $\beta = .02$ ,  $p = .008$ ,  $CI = [.01, .07]$ ), enthusiasm ( $\beta = .02$ ,  $p = .006$ ,  $CI = [.01, .05]$ ), and hostile cognitions ( $\beta = .01$ ,  $p = .020$ ,  $CI = [.00, .03]$ ) through the path of motive attributions. However, preceding uncivil comments did not show such an effect (anger:  $\beta = .00$ ,  $p = .922$ ,  $CI = [-.01, .04]$ , anxiety:  $\beta = .00$ ,  $p = .922$ ,  $CI = [-.02, .06]$ , enthusiasm:  $\beta = .00$ ,  $p = .903$ ,  $CI = [-.02, .05]$ , and hostile cognitions:  $\beta = .00$ ,  $p = .835$ ,  $CI = [-.01, .01]$ ). Further, data revealed that anger ( $\beta = .24$ ,  $p < .001$ ), enthusiasm ( $\beta = .08$ ,  $p = .048$ ), and hostile cognitions ( $\beta = .09$ ,  $p = .022$ ) positively influenced intentions to answer in an aggressive manner whereas anxiety did not influence this variable ( $\beta = .00$ ,  $p = .916$ ). Moreover, anger ( $\beta = .22$ ,  $p < .001$ ) and enthusiasm ( $\beta = .23$ ,  $p < .001$ ) positively influenced individuals' intentions to answer in an discussion-

centered way. In contrast, anxiety ( $\beta = -.00$ ,  $p = .939$ ) and hostile cognitions ( $\beta = .01$ ,  $p = .771$ ) did not affect this intention.

We also expected that individuals are more likely to write an uncivil answer to the basic uncivil comment (H4a) and to criticize its sender (H4b) when attributing higher levels of aggressive motives to the sender. Although not hypothesized a priori, based on well-established psychological mechanisms (Ajzen, 1991), it was also tested whether individuals' aggressive answering intentions functions as a mediator within this relationship. Therefore, for each dependent variable, a mediation analysis was conducted. Please note that results for the dependent variables are reported as log-odds metrics.

In both models, aggressive motive attributions increased aggressive answering intentions message ( $R^2 = .09$ ,  $F(1, 609) = 57.96$ ,  $p < .001$ ;  $B = .49$ ,  $SE = .06$ ,  $p < .001$ ).

When examining the effect of aggressive motive attributions on uncivil answers, the mediation model was significant ( $\chi(1) = 319.52$ ,  $p < .001$ , Nagelkerke's  $R^2 = .15$ ). Against H4a, aggressive motive attributions did not affect the likelihood of using uncivil expressions ( $B = -.04$ ,  $SE = .02$ ,  $Z = -1.72$ ,  $p = .086$ ,  $CI = [-.08, .01]$ ). However, participants aggressive answering intentions positively affected uncivil answering behavior ( $B = .07$ ,  $SE = .01$ ,  $Z = 6.28$ ,  $p < .001$ ,  $CI = [.05, .10]$ ). Although the direct effect of aggressive motive attributions was not significant, the indirect effect through aggressive answering intentions was significant – albeit small ( $B = .03$ ,  $SE = .01$ ,  $CI = [.02, .05]$ ). When entering critique expressions as dependent variable, the overall model was significant ( $\chi(1) = 791.40$ ,  $p < .001$ , Nagelkerke's  $R^2 = .05$ ). Noteworthy, the model was not well-fitting as indicated by Nagelkerke's  $R^2$ . As expected (H4b), aggressive motive attributions increased the likelihood of criticizing the sender of the basic uncivil message ( $B = .06$ ,  $SE = .01$ ,  $Z = 4.26$ ,  $p < .001$ ,  $CI = [.03, .08]$ ). Likewise, aggressive answering intentions significantly, but negatively, affected the likelihood of critique expressions ( $B = -.02$ ,  $SE = .01$ ,  $Z = -3.21$ ,  $p = .001$ ,  $CI =$

[-.04, -.01]). Consequently a significant but negative indirect effect emerged ( $B = -.02$ ,  $SE = .00$ ,  $CI = [-.02, -.00]$ ).

### **Discussion**

By drawing on the general aggression model (Anderson & Bushman, 2002), the current study examined how people process an uncivil comment based on situational factors. We expected that the extent of incivility in comments preceding an uncivil comment and the stance of an uncivil comment influence individuals' internal state variables (i.e., emotions and cognitions) as well as the extent of aggressive attribution making. In turn, we expected that this process directs recipients' reactions to the comment. Although our theoretical model did not fit the empirical data, results from an adapted path model that still corresponds to the theoretical framework (Anderson & Bushman, 2002) provides valuable insights into how an uncivil comment affects its recipients. We found that the uncivil comment's stance is more important than the extent of incivility in preceding comments. However, the direct effects of the experimental conditions on individuals' emotions and cognitions (H1.1a – d & H2.1a – d) could not be examined as the alternative model did not include these paths. Accordingly, the indirect effects of the experimental conditions on the attribution of aggressive motives through the pathway of the internal state variables (H1.3a–d & H2.3a–d) could not be investigated. Nevertheless, against our expectations (H1.2), people were not more likely to attribute aggressive motives to the sender if an uncivil comment was accompanied by preceding uncivil comments instead of civil comments. However, as proposed (H2.2), people were more likely to attribute aggressive motives to the sender of an uncivil comment when the comment opposed the recipient's stance. Contradicting our assumptions (H3a), individuals were not less likely to answer in a discussion-centered manner to the basic uncivil comment when attributing higher levels of aggressive motives to the sender. Supporting H3b, they were more likely to answer aggressively. Moreover, as expected (H4a), individuals were

slightly more likely to criticize the sender of the uncivil message but did not use more incivility in their answers as proposed by H4b. However, mediation analyses provide a more comprehensive picture: It was found that aggressive motive attributions indirectly increased the likelihood of uncivil comment behavior, whereas critique behavior was decreased through this path.

Our results suggest that people appraise individuals' motives for behaving uncivilly independent of others' uncivil commenting behavior. Compared to preceding civil comments, neither preceding uncivil comments nor the absence of preceding comments influenced individuals' attribution of aggressive motives to the sender of an uncivil comment. This result can have many potential reasons. For instance, the perception of descriptive norms, as proposed by Cialdini and colleagues (1990), might not have been strong enough to outweigh perceptions of injunctive norms. Thus, people still considered incivility as unacceptable behavior in discussions. Indeed, some comments that were written by participants indicated that they considered the uncivil basic comment to be an adaptive behavior in the condition with preceding uncivil comments but still disapproved of this kind of communication. For instance, one person commented: "Please remain objective and try to understand the motives of the others. You should not stoop to the level of others." Another person stated: "You may have been tempted by the rude and impolite language of the other posts in this forum, but should we not discuss such a serious topic objectively?" In this vein, the initial Facebook post to which the manipulated comments answered were written in a civil manner. This may have been enough to activate an injunctive norm, as it reminded individuals that incivility is usually disapproved (see also Cialdini et al., 1990). Another explanation could be that the presentation of uncivil user comments at one time was not sufficient to activate a descriptive norm that mitigates the effects of incivility. A relatively high number of participants (32.9%) stated that they commented on social media discussions at least once a week, and the use of

social media was prevalent. Therefore, in terms of the GAM (Anderson & Bushman, 2002), people might have developed more stable knowledge structures regarding incivility through experiences that cannot be easily altered. Moreover, important to note, in the current study, the preceding comments and the basic uncivil comment represented the same stance. Thus, when stances in uncivil discussions are mixed, people might make different appraisals.

In accordance with our expectations and previous findings (Reeder et al., 2005), people attributed more aggressive motives to the sender of an uncivil comment when it opposed their stance (compared to a like-minded stance). As stated by the Self-Categorization Theory (Turner et al., 1987), individuals probably perceived persons with opposing stances as members of a competing group. Nevertheless, we could not confirm that this tendency is mediated by individuals' negative emotions. As argued earlier, we assumably did not capture individuals' initial rudimentary emotions. It is, therefore, still plausible that people attribute negative motives to opposing stances to mitigate negative feelings triggered by cognitive dissonance (Cooper & Mackie, 1983).

Regarding our research question (RQ1), no interaction effect of preceding comments and the stance on aggressive motive attributions has been found. As the type of preceding comments did not influence these attributions, it is not surprising that no interaction effect emerged. However, the influence of the comment's stance on the perception of incivility can be very complex. Even the type of incivility (e.g., insult vs. vulgarity) might determine how it is perceived (Gervais, 2017). Thus, future research should not underestimate the influence of the comment's stance on the perception of incivility.

Further, aggressive attributions significantly increased individuals' anger, anxiety, enthusiasm, and hostile cognitions. Although not hypothesized in this way, these results provide valuable insights into the mechanisms of how incivility affects individuals. Moreover, aggressive motive attributions directly increased intentions to answer

aggressively. However, this tendency was in part mediated by internal state variables. Similarly, aggressive motive attributions only indirectly increased intentions to answer in a discussion-centered manner indicating a full mediation effect through anger and enthusiasm. In this respect, it was already demonstrated that incivility can increase anger (e.g., Chen et al., 2020), anxiety (Lu & Myrick, 2016), and hostile cognitions (Rösner et al., 2016). As aggressive motive attributions also increased these variables, this suggests that individuals' appraisal of an uncivil comment can reinforce or alter rudimentary emotions and cognitions triggered by exposure to an uncivil comment. Moreover, anger and hostile cognitions in turn affected individuals' intentions to answer aggressively to the basic uncivil comment.

Additionally and surprisingly, anger also increased individuals' intentions to answer in a discussion-centered way. The effect of aggressive motive attributions, as well as the effects of anger and hostile cognitions on aggressive answering intentions, correspond to the assumptions of the GAM (Anderson & Bushman, 2002), that an aversive internal state and a negative appraisal increase the likelihood of aggressive reactions. However, anger also increased the likelihood of constructive behavior. This result is particularly important as it shows that anger can produce both constructive and aggressive reactions. This might also explain why incivility can simultaneously have positive and negative effects (e.g., Borah, 2014). Importantly, the two dimensions of answering intentions (discussion-centered or aggressive) are not mutually exclusive. For instance, people can have the intent to show uncivil others that they are "idiots" but still are motivated to understand those persons better. In this respect, anger might serve as a motivating force to "neutralizing disliked elements" (Marcus et al., 2006, p. 36). To get in a discussion-centered conversation with others might simply be a strategy to avoid further incivility.

This assumption is supported by our finding that aggressive motive attributions increased the likelihood of criticizing the sender and not the likelihood of behaving uncivilly

– albeit this effect was small. However, it was also found that the attribution of aggressive discussion motives can indirectly increase the likelihood of uncivil commenting behavior and decrease the likelihood of critique expressions through the pathway of aggressive answering intentions. Therefore, when aggressive motive attributions increase aggressive discussion intentions, people seem to be more likely to engage in uncivil commenting behavior and less likely to criticize others.

Contradicting earlier findings regarding the impact of incivility (cf. Hutchens et al., 2019), aggressive motive attributions increased individuals' feelings of enthusiasm, too. In the context of political stimuli, enthusiasm is associated with positive goal pursuit, whereas feelings like anger are associated with neutralizing undesired inputs (Marcus et al., 2000). Otherwise, these emotional states can probably co-occur. Concerning the current study, people might have felt increased anger as they perceived an undesired threat which is reflected by attributions of aggressive motive attributions. At the same time, they can also be convinced that their behavior is much more purposeful which in turn triggers feelings of enthusiasm. Indeed, enthusiasm positively influenced individuals' intentions to answer to the basic uncivil comment in a discussion-centered manner. This also explained why aggressive motive attributions indirectly increased participants' intention to answer in a discussion-based manner. Noteworthy, this was also true for the aggressive answering intention, but the effect was relatively small. Therefore, being confronted with a threat might lead to aggressive motive attributions but if individuals are convinced that their behavior is more purposeful, this probably encourages them to further action in this way – i.e., constructively. Thus, aggressive motive attributions only indirectly increased the likelihood of reacting in a discussion-centered manner.

Interestingly, in the study's context, anxiety did not affect individuals' behavioral intentions. Indeed, although associations between incivility and anxiety have been found

(Gervais, 2017; Lu & Myrick, 2016), there is only limited evidence that anxiety directly affects individuals' behavior. In general, anxiety is connected to a surveillance system that directs individuals' attention (Marcus et al., 2000, 2006). Therefore, anxiety might rather drive people to take a closer look at the situation but not directly to alter their behavior. Future studies should investigate which concrete role anxiety plays in online discussions. It must be noted that only a small amount of the variances of all internal state variables could be explained (1 – 4%). Thus, there are assumably other variables that guide individuals' emotional states.

### **Limitations and Further Directions**

Some limitations of the study have to be acknowledged. First, the representativeness of the sample is limited as we draw on a convenience sample. Further, participants' self-reported emotions and cognitions were measured with a time delay. Therefore, we probably did not capture their initial immediate affective state, which led to the poor fit of the original theoretical model. Future studies should measure the internal state more implicitly and immediately (e.g., by physical measurements). Moreover, the stance of preceding comments and the basic uncivil comment were kept constant. Future studies should investigate how discussions with mixed stances influence the appraisal of uncivil comments. Further, we did not measure all motives which people can attribute to the sender of an uncivil message and did not measure all possible answering intentions. Likewise, many forms of incivility can affect individuals differently. Thus, it could be examined how different types of incivility alter the attribution of motives. Finally, participants were aware that they were exposed to a hypothetical scenario. Nevertheless, this does not necessarily mitigate the effects of incivility, as the relevant knowledge structures are assumably activated to a similar extent.

## **Conclusion**

This study provides insights into which ways an uncivil comment is processed depending on accompanied (uncivil) comments and its represented stance. Surprisingly, whether an uncivil comment is posted in civil or uncivil discussions seems to be less influential. More important is that people tend to attribute more aggressive motives to the sender of an opposing uncivil comment than to the sender of a like-minded uncivil comment. This indicates that in cross-cutting discussions civility might be of particular importance to mitigate undesired additional detrimental effects on attribution processes. Moreover, we found that the interplay of individuals' emotions and cognitions and the appraisal of an uncivil comment significantly impact individuals' intentions to react. However, while anger fostered constructive and aggressive answering intentions to a similar extent, the attribution of aggressive discussion motives increased, in particular aggressive answering intentions. Thus, future interventions should preliminarily focus on how attributions can be altered to avoid the negative effects of incivility.

## References

- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179–211. [https://doi.org/10.1016/0749-5978\(91\)90020-T](https://doi.org/10.1016/0749-5978(91)90020-T)
- Allen, J. J., Anderson, C. A., & Bushman, B. J. (2018). The General Aggression Model. *Current Opinion in Psychology*, 19, 75–80. <https://doi.org/10.1016/j.copsyc.2017.03.034>
- 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., & Bushman, B. J. (2002). Human Aggression. *Annual Review of Psychology*, 53(1), 27–51. <https://doi.org/10.1146/annurev.psych.53.100901.135231>
- Berkowitz, L. (1990). On the formation and regulation of anger and aggression: A cognitive-neoassociationistic analysis. *American Psychologist*, 45(4), 494–503. <https://doi.org/10.1037/0003-066X.45.4.494>
- Bicchieri, C. (2017). *Norms in the wild: How to diagnose, measure, and change social norms*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780190622046.001.0001>
- 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>
- Bushman, B. J., & Anderson, C. A. (2002). Violent video games and hostile expectations: A test of the general aggression model. *Personality and Social Psychology Bulletin*, 28(12), 1679–1686. <https://doi.org/10.1177/014616702237649>
- Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2nd ed.). Routledge Taylor & Francis Group.

<https://doi.org/10.4324/9780203805534>

- 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., Lu, S., & Fadnis, D. (2020). Does online incivility cancel out the spiral of silence? A moderated mediation model of willingness to speak out. *New Media & Society*. Advance online publication. <https://doi.org/10.1177/1461444820954194>
- Chen, G. M., Muddiman, A., Wilner, T., Pariser, E., & Stroud, N. J. (2019). We Should Not Get Rid of Incivility Online. *Social Media + Society*, 5(3). <https://doi.org/10.1177/2056305119862641>
- Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology*, 58(6), 1015–1026. <https://doi.org/10.1037/0022-3514.58.6.1015>
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4), 284–290. <https://doi.org/10.1037/1040-3590.6.4.284>
- Clore, G. L., & Centerbar, D. B. (2004). Analyzing anger: How to make people mad. *Emotion*, 4(2), 139–144. <https://doi.org/10.1037/1528-3542.4.2.139>
- 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>
- Cooper, J., & Mackie, D. (1983). Cognitive dissonance in an intergroup context. *Journal of Personality and Social Psychology*, 44(3), 536–544. <https://doi.org/10.1037/0022-3514.44.3.536>

- Gervais, B. T. (2015). Incivility Online: Affective and Behavioral Reactions to Uncivil Political Posts in a Web-based Experiment. *Journal of Information Technology and 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>
- 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* (2nd ed.). Guilford. <https://doi.org/978-1-60918-230-4>
- Hayes, A. F., & Montoya, A. K. (2017). A Tutorial on Testing, Visualizing, and Probing an Interaction Involving a Multicategorical Variable in Linear Regression Analysis. *Communication Methods and Measures*, 11(1), 1–30. <https://doi.org/10.1080/19312458.2016.1271116>
- Hogg, M. A., & Reid, S. A. (2006). Social identity, self-categorization, and the communication of group norms. *Communication Theory*, 16(1), 7–30. <https://doi.org/10.1111/j.1468-2885.2006.00003.x>
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(2), 179–185. <https://doi.org/10.1007/BF02289447>
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Hutchens, M. J., Silva, D. E., Hmielowski, J. D., & Cicchirillo, V. J. (2019). What's in a username? Civility, group identification, and norms. *Journal of Information Technology*

& *Politics*, 16(3), 203–218. <https://doi.org/10.1080/19331681.2019.1633983>

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>

Hyun, K. D., & Kim, J. (2015). Differential and interactive influences on political participation by different types of news activities and political conversation through social media. *Computers in Human Behavior*, 45, 328–334. <https://doi.org/10.1016/j.chb.2014.12.031>

Jones, E. E., & Davis, K. E. (1965). From Acts To Dispositions The Attribution Process in Person Perception. *Advances in Experimental Social Psychology*, 2, 219–266. [https://doi.org/10.1016/S0065-2601\(08\)60107-0](https://doi.org/10.1016/S0065-2601(08)60107-0)

Lu, Y., & Myrick, J. G. (2016). Cross-cutting exposure on Facebook and political participation: Unraveling the effects of emotional responses and online incivility. *Journal of Media Psychology*, 28(3), 100–110. <https://doi.org/10.1027/1864-1105/a000203>

Malle, B. F. (2011). Attribution Theories: How people make sense of behavior. In D. Chadee (Ed.), *Theories in social psychology* (pp. 72–95). Wiley Blackwell.

Marcus, G. E., MacKuen, M., Wolak, J., & Keele, L. (2006). The Measure and Mismeasure of Emotion. In D. P. Redlawsk (Ed.), *Feeling politics: Emotion in political information processing* (pp. 31–45). Palgrave Macmillan. <https://doi.org/10.1017/CBO9781107415324.004>

Marcus, G. E., Neuman, W. R., & MacKuen, M. (2000). *Affective Intelligence and Political Judgment*. University of Chicago Press.

Reeder, G. D. (2009). Mindreading: Judgments About Intentionality and Motives in Dispositional Inference. *Psychological Inquiry*, 20(1), 1–18.

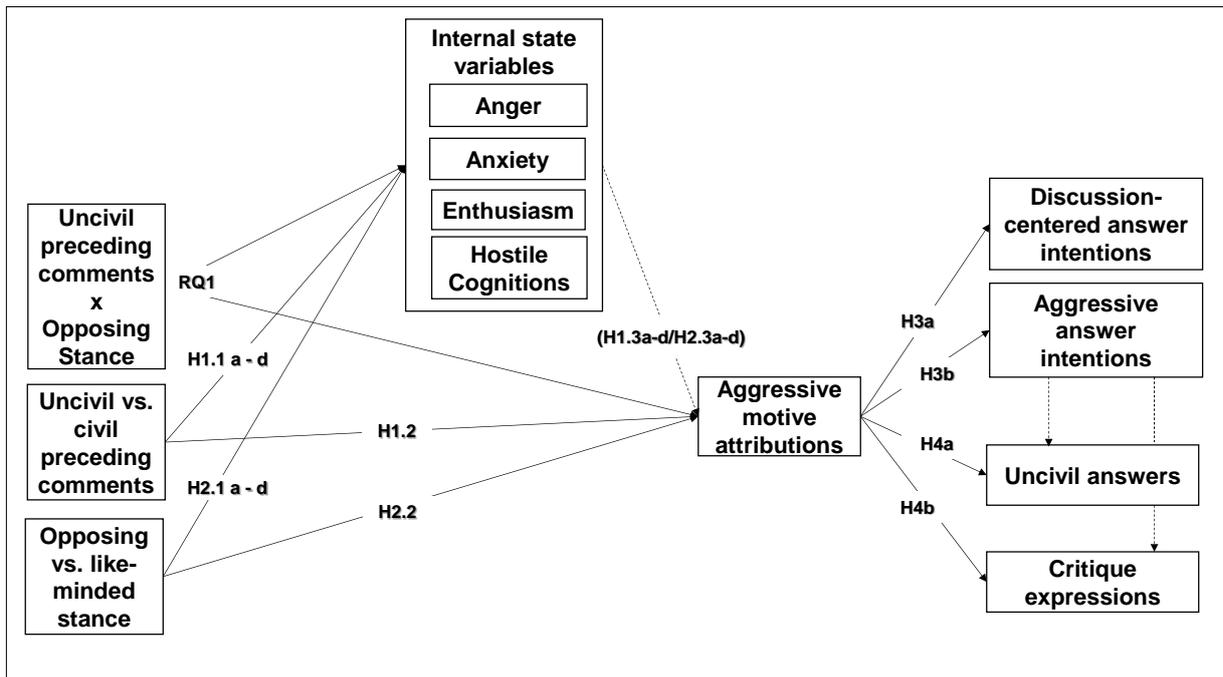
<https://doi.org/10.1080/10478400802615744>

- Reeder, G. D., Kumar, S., Hesson-McInnis, M. S., & Trafimow, D. (2002). Inferences about the morality of an aggressor: The role of perceived motive. *Journal of Personality and Social Psychology*, *83*(4), 789–803. <https://doi.org/10.1037/0022-3514.83.4.789>
- Reeder, G. D., Pryor, J. B., Wohl, M. J. A., & Griswell, M. L. (2005). On attributing negative motives to others who disagree with our opinions. *Personality and Social Psychology Bulletin*, *31*(11), 1498–1510. <https://doi.org/10.1177/0146167205277093>
- Reeder, G. D., & Trafimow, D. (2004). Attributing Motives to other People. In *Other minds: How humans bridge the divide between self and others* (pp. 98–115). Guilford Press.
- Rösner, L., & Krämer, N. C. (2016). Verbal Venting in the Social Web: Effects of Anonymity and Group Norms on Aggressive Language Use in Online Comments. *Social Media + Society*, *2*(3). <https://doi.org/10.1177/2056305116664220>
- 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>
- 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 and Society*, *20*(10), 3678–3699.  
<https://doi.org/10.1177/1461444818757205>
- Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Margaret, S. (1987). Rediscovering the Social Group: A Self-Categorization Theory. In J. C. Turner, M. A. Hogg, P. J. Oakes, S. S. Reicher, & M. S. Wetherell (Eds.), *Rediscovering the Social*

*Group: A Self-Categorization Theory*. Basil Blackwell.

Wang, M. Y., & Silva, D. E. (2017). 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>



**Figure 1.** Conceptual model visualizing H1.1a – H4b. Dashed lines represent indirect paths. Hypotheses in parentheses represent the indirect effects of the experimental conditions on aggressive motive attributions.

Table 1

*Examples of comments employed in the study.*

	Pro Issue ( $n = 310$ )	Con Issue ( $n = 301$ )
Basic uncivil comment to be appraised	Yeah man!! We all have something of that... But well, there are apparently always a few dumb asses who think that meat belongs in school. But they should better think about whether they can't keep their shitty attitude to themselves!	No man!! Nobody has anything of that... But well, there are apparently always a few dumb asses who think that meat doesn't belong in school at all. But they should better think about whether they can't keep their shitty attitude to themselves!
Uncivil preceding comment	I think that shitty politics should generally pay more attention to the demands for a meat ban in schools... It makes me want to puke if some fools want to tell me that this is not the wish of the majority.	I think that shitty politics should generally pay less attention to the demands for a meat ban in schools... It makes me want to puke if some fools want to tell me that this is the wish of the majority.
Civil preceding comment	I think that politics should generally pay more attention to the demand for a meat ban in schools... Even though I respect other opinions, I believe that this is the wish of the majority.	I think that politics should generally pay less attention to the demand for a meat ban in schools... Even though I respect other opinions, I do not believe that this is the wish of the majority.

*Note.* Example comments are only presented for the topic of meat ban in school canteens.

Table 2  
*Bivariate correlations of variables in the path model.*

	(1)	(2)	(3)	(4)	(5)	(6)
1. Hostile Cognitions						
2. Enthusiasm	-.04					
3. Anger	.23***	.04				
4. Anxiety	.07	.10*	.68***			
5. Aggressive Attributions	.12**	.19***	.11**	.19***		
6. Discourse Intention	.05	.23***	.23***	.16***	.03	
7. Aggressive Intention	.17***	.13**	.30***	.23***	.30***	.29***

\* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$

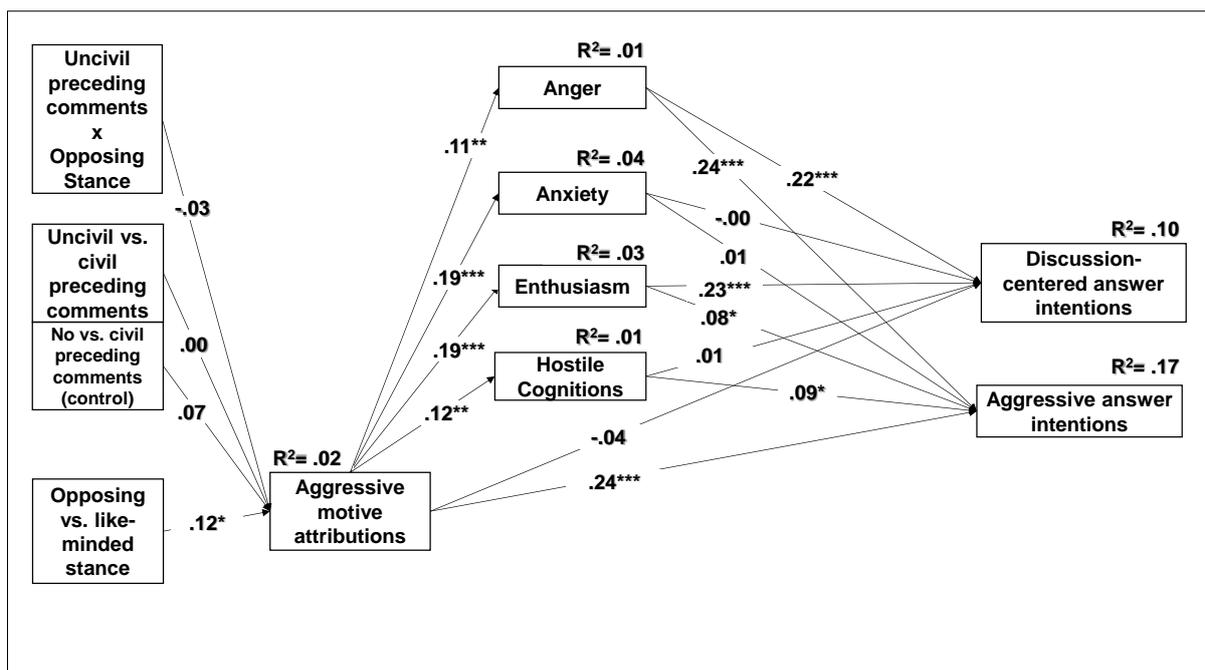


Figure 2. Computed path model. Standardized effects are reported.

\* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$

## Appendix A: Pre-Test

Within the pre-test, the six comments intended for the main study, as well as three extreme uncivil and three civil comments have been evaluated by participants ( $N = 30$ ). In random order, each statement was rated on a 7-point semantic differential, including four adjective pairs reflecting perceived incivility (*polite – impolite*, *respectful – disrespectful*, *friendly – hostile*, and *peaceful – aggressive*). No discussion topic was mentioned in the pre-test. A repeated-measures ANOVA<sup>4</sup> revealed that there was a significant difference between comments regarding the perceived incivility ( $F(4.931, 143.003) = 304.84, p < .001, \eta^2 = .913$ ). Pairwise comparisons with Bonferroni corrections revealed that the six comments intended for the study did not differ significantly among each other (each pair  $p > .05$ ). However, the six comments differed significantly from the three civil comments (each pair  $p < .001$ ) and the three extreme uncivil comments (each pair  $p < .05$ ). See *Table 1* for means and standard deviations.

Table 1

*Means and standard deviations of perceived incivility.*

Comment Type	<i>M</i>	<i>SD</i>
<i>Comments for the main study</i>		
Uncivil Comment 1	22.83	0.67
Uncivil Comment 2	22.17	0.73
Uncivil Comment 3	24.67	0.50
Uncivil Comment 4	23.40	0.69
Uncivil Comment 5	23.67	0.61
Basic Uncivil Comment	23.73	0.65
<i>Comparative comments</i>		
Extreme uncivil comment 1	27.77	0.17
Extreme uncivil comment 2	27.90	0.07
Extreme uncivil comment 3	26.67	0.39
Civil comment 1	5.63	0.68
Civil comment 2	6.60	0.67
Civil comment 3	5.40	0.43

<sup>4</sup> As Mauchly's Tests of sphericity has been significant, Greenhouse-Geisser corrections are reported.

## Appendix B: Results of the Exploratory Factor Analyses

Table 1

*Results of the EFAs with principal axis analysis and promax rotation for the measures of hostile attributions and intentions to react.*

Items	<i>Factor 1</i>	<i>Factor 2</i>
<i>Hostile attribution</i>		
...to unsettle other people	.72	
...to threaten other people	.70	
...to annoy other people	.69	
...to attack other people	.65	
...to disturb the discussion	.59	
...to promote the discussion (reserved)		.73
...to encourage other people (reserved)		.59
<i>Intentions to react</i>		
...to signal the person to shut up	.86	
...to show the person that she/he is an idiot	.84	
...to express anger about the person	.82	
...to show the person that she/he is inferior	.80	
...to humiliate the person	.80	
...to pay back the person	.78	
...to punish the person in the discussion	.76	
...to tell the person to leave the discussion	.75	
...to better understand the opinion of the person		.81
...to learn more about the person's opinion		.81
...to discuss the topic with the person respectfully		.77
...to get into a deeper conversation with the person		.73
...to share her/his knowledge with the person		.64
...to motivate the person to continue expressing their opinion		.64
...to show that nobody has anything to fear from other opinions		.59

*Note.* Bartlett's test of sphericity for hostile attributions:  $\chi^2 = 860.28$ ,  $df = 10$ ,  $p < .001$  and for intentions to react:  $\chi^2 = 6093.06$ ,  $df = 105$ ,  $p < .001$ .

### Appendix C: Moderation Analyses

In the following, results of the moderation analyses as hypothesized in the preregistration are presented. The hypotheses were tested by using the PROCESS macro for SPSS (Model 1; Hayes, 2018). Uncivil preceding comments and the control with no preceding comments were tested against the civil preceding comment condition.

Results of the moderation analyses H5a-H7d are presented in *Table 1*. It was stated that the effects of preceding uncivil comments on individuals' emotions are altered by their levels of psychopathy (H5a-c) and Machiavellianism (H6a-c). Testing the moderating effects of psychopathy, results revealed that when using anger as the dependent variable, the overall model was significant ( $F(5,605) = 10.04, p < .001, R^2 = .08$ ). However, neither preceding uncivil comments nor the control condition affected anger. Only psychopathy had a positive effect, but no significant interaction with preceding uncivil comments nor the control emerged. When using anxiety as dependent variable, the overall model indicated significance, too ( $F(5,605) = 2.25, p = .048, R^2 = .02$ ). However, again, psychopathy was the only significant predictor, and no interaction effect emerged. For enthusiasm, the overall model was not significant ( $F(5,605) = 1.58, p = .163, R^2 = .01$ ). Thus, H5a-c were not confirmed.

Testing the effects of Machiavellianism as moderator, the overall model for anger as dependent variable was significant ( $F(5,605) = 9.61, p < .001, R^2 = .07$ ). However, only the moderator predicted anger but neither the experimental conditions nor the interaction terms did. The overall model for anxiety was not significant ( $F(5,605) = 0.76, p = .577, R^2 = .01$ ). None of the included variables predicted anxiety. When entering enthusiasm as dependent variable, the overall model was significant ( $F(5,605) = 2.97, p = .012, R^2 = .02$ ). Data revealed that although only the control had a direct positive effect on enthusiasm, the interaction of uncivil preceding comments and Machiavellianism was significant, too. 1% of explained variance increased statistically significant ( $F(2,605) = 4.22, p = .015, \Delta R^2 = .01$ ). Conditional effects showed that there was a significant effect of uncivil preceding comments on enthusiasm for low Machiavellianism (-1 SD;  $B = 3.64, SE = 1.53, p = .018, CI = [.63, 6.65]$ ) but not for average ( $B = 1.53, SE = 1.03, p = .138, CI = [-.49, 3.55]$ ) or high Machiavellianism (+1 SD;  $B = -.59, SE = 1.44, p = .684, CI = [-3.41, 2.24]$ ). Thus, H6a&b had to be rejected whereas H6c was confirmed.

Moreover, H7a-d proposed that narcissism moderates the effects between the exposure to opposing stances (vs. like-minded stances) and emotions as well as hostile cognitions. When adding anger as dependent variable, the overall model reached significance ( $F(3,607) = 2.97, p = .004, R^2 = .02$ ). However, only narcissism influenced individuals' feelings of

anger. Moreover, neither anxiety ( $F(3,607) = 0.74, p = .530, R^2 = .00$ ), enthusiasm ( $F(3,607) = 0.28, p = .843, R^2 = .00$ ), nor hostile cognitions ( $F(3,607) = 0.45, p = .718, R^2 = .00$ ) were predicted by an opposing stance, the moderator or the interaction terms.

Table 1

*Effects of variables included in the moderation analyses to examine H5a – H7d.*

	Anger <i>B (SE)</i> [95% CI]	Anxiety <i>B (SE)</i> [95% CI]	Enthusiasm <i>B (SE)</i> [95% CI]	Hostile cognitions <i>B (SE)</i> [95% CI]
<b>H5a-c</b>				
Preceding uncivil comments	-.12 (.73)	.56 (.82)	1.48 (1.03)	
Control	.28 (.73)	.03 (.83)	2.33 (1.04)*	
Psychopathy	.26 (.08)***	.18 (.09)*	-.04 (.11)	
Preceding uncivil comments x Psychopathy	-.02 (.11) [-.23, .20]	-.06 (.12) [-.21, -.05]	-.06 (.15) [-.36, .24]	
Control x Psychopathy	.13 (.11) [-.08, .34]	-.01 (.12) [-.30, .18]	.18 (.15) [-.12, .48]	
<b>H6a-c</b>				
Preceding uncivil comments	.36 (.73)	.77 (.83)	1.53 (1.03)	
Control	.80 (.74)	.20 (.84)	2.60 (1.04)*	
Machiavellianism	.22 (.06)***	.10 (.07)	.10 (.09)	
Preceding uncivil comments x Machiavellianism	-.07 (.08) [-.24, .09]	-.05 (.09) [-.24, .13]	-.23 (.12)* [-.45, -.00]	
Control x Machiavellianism	.05 (.08) [-.11, .20]	-.06 (.09) [-.24, .11]	.07 (.11) [-.15, .30]	
<b>H7a-d</b>				
Opposing stance	-.25 (.61)	-.07 (.68)	-.16 (.85)	-.73 (.63)
Narcissism	.16 (.05)***	.08 (.05)	.05 (.07)	.09 (.10)
Opposing stance x Narcissism	-.12 (.09) [-.31, .06]	-.02 (.10) [-.23, .18]	-.03 (.13) [-.29, .22]	-.06 (.19) [-.43, .32]

Finally, we predicted that individuals high in trait emotional intelligence are less likely to have aggressive answer intentions when attributing hostile motives to the sender of a message (H8). However, the hypothesis was not confirmed. Although the overall model was significant ( $F(3,607) = 37.68, p < .001, R^2 = .16$ ) and both hostile motive attributions ( $B = .48, SE = 0.06, p < .001$ ) as well as trait emotional intelligence ( $B = -.14, SE = 0.02, p < .001$ ) affected individuals' aggressive answer intentions, the interaction yielded no significance ( $B = -.00, SE = .00, p = .459, CI = [-.01, .00]$ ).

## **Reference**

Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis* (2nd ed.). Guilford. <https://doi.org/978-1-60918-230-4>

# DuEPublico

Duisburg-Essen Publications online

UNIVERSITÄT  
DUISBURG  
ESSEN

*Offen im Denken*

ub | universitäts  
bibliothek

Diese Dissertation wird via DuEPublico, dem Dokumenten- und Publikationsserver der Universität Duisburg-Essen, zur Verfügung gestellt und liegt auch als Print-Version vor.

**DOI:** 10.17185/duepublico/75896  
**URN:** urn:nbn:de:hbz:465-20220519-083349-2

Alle Rechte vorbehalten.