

Wandel der Sprach- und Debattenkultur in sozialen Online-Medien

**Wandel der Sprach- und Debattenkultur
in sozialen Online-Medien**

Ein Literaturüberblick zu Ursachen und Wirkungen
von inziviler Kommunikation

Anna Sophie Kümpel, Diana Rieger

Impressum

Herausgeberin:

Konrad-Adenauer-Stiftung e. V., 2019, Berlin

Ansprechpartner:

Konrad-Adenauer-Stiftung e. V.

Thomas Köhler

Leiter der Hauptabteilung Politik und Beratung

Hauptabteilung Politik und Beratung

T: +49 30 / 26 996-3550

thomas.koehler@kas.de

Sabine Stoye

Projektkoordinatorin Wandel der Sprach- und Debattenkultur

Hauptabteilung Politik und Beratung

T: +49 30 / 26 996-3517

sabine.stoye@kas.de

Daphne Wolter

Koordinatorin Medienpolitik

Hauptabteilung Politik und Beratung

T: +49 30 / 26 996-3607

daphne.wolter@kas.de

Gestaltung: yellow too Pasiak Horntrich GbR

Satz: Janine Höhle, Konrad-Adenauer-Stiftung e. V.

Die Printausgabe wurde bei der Druckerei copy print Kopie & Druck GmbH, Berlin gedruckt.

Printed in Germany.

Gedruckt mit finanzieller Unterstützung der Bundesrepublik Deutschland.



Der Text dieser Publikation ist lizenziert unter den Bedingungen von „Creative Commons Namensnennung-Weitergabe unter gleichen Bedingungen 4.0 international“, CC BY-SA 4.0 (abrufbar unter: <https://creativecommons.org/licenses/by-sa/4.0/legalcode.de>).

ISBN 978-3-95721-546-8

Wandel der Sprach- und Debattenkultur in sozialen Online-Medien

Ein Literaturüberblick zu Ursachen und Wirkungen
von inziviler Kommunikation

Anna Sophie Kümpel, Diana Rieger

Inhaltsverzeichnis

1. Einleitung	5
2. Begriffsklärungen und Perspektive des Beitrags	9
3. Ursachen für die Wahrnehmung von mehr Inzivilität in sozialen Online-Medien	13
4. Wirkungen von Inzivilität in sozialen Online-Medien	19
5. Gegenmaßnahmen: Wie lässt sich Inzivilität in sozialen Online-Medien verhindern?	25
6. Fazit und Ausblick	31
Literaturverzeichnis	35

1. Einleitung

1. Einleitung

Online-Medien sind zu einem selbstverständlichen Bestandteil unseres Alltags geworden. Gemäß der ARD/ZDF-Onlinestudie sind mehr als 90 Prozent der Deutschen online, mehr als die Hälfte von ihnen nutzt das Internet täglich (Frees & Koch, 2018). Insbesondere soziale Online-Medien spielen für viele Nutzer dabei eine wichtige Rolle und wurden 2018 von mehr als 30 Prozent der Gesamtbevölkerung mindestens einmal wöchentlich genutzt (ebd.).

Während insbesondere soziale Netzwerkseiten (SNS) wie Facebook und Twitter anfänglich vor allem der interpersonalen Vernetzung und Selbstdarstellung dienen, sind die Plattformen indes auch zu einer zentralen Schnittstelle für die Medien-, Informations- und Nachrichtennutzung geworden (Hölig & Hasebrink, 2018; Kümpel, 2019). Angesichts der vielfältigen Potentiale für Partizipation, Interaktion und Anschlusskommunikation haben sie zudem dazu beigetragen, dass a) die Teilhabe an öffentlichen Diskursen drastisch vereinfacht wurde, b) Nutzer*innen flexibel zwischen Kommunikator- und Rezipientenrolle wechseln können und c) journalistische Inhalte sowie politische Debatten auch von „normalen“ Bürgern öffentlichkeitswirksam kommentiert und problematisiert werden können (Neuberger, 2017). Dies gilt auch für weitere Formen sozialer Online-Medien wie Foren und Online-*Communities*, Kommentarsektionen auf Nachrichten-Webseiten sowie Videoplattformen und Wikis, denen wir uns im Rahmen dieses Beitrags widmen werden.

Indem sie einen öffentlichen Raum bereitstellen, in welchem Nutzer*innen ihre eigenen Meinungen, Perspektiven und Fachkenntnisse einbringen können, wurde sozialen Online-Medien gerade in ihrer Anfangsphase ein hohes deliberatives Potenzial zugeschrieben (Rowe, 2015b; Ruiz u. a., 2011). Viele Beobachter*innen sahen nicht zuletzt in der Anonymität des Netzes die Möglichkeit, soziale Unterschiede zu nivellieren, einen produktiven gesellschaftlichen Austausch zu fördern und so den

Weg für eine „democratic utopia“ (Papacharissi, 2004, S. 260) zu ebnet (siehe auch Chen, 2017, S. 38 ff.). Tatsächlich scheinen Nutzer*innen sozialer Online-Medien im Vergleich zu *Face-to-Face-Settings* stärker gewillt, ihre eigene Meinung zu vertreten und sich auch über sensible Themen auszutauschen (Ho & McLeod, 2008). Dies wiederum kann sich positiv auf das Kompetenz- und Autonomieerleben sowie den Zusammenhalt in (virtuellen) Gruppen auswirken (für einen Überblick siehe Christopherson, 2007).

Die veränderten Kommunikationsbedingungen ziehen jedoch fraglos nicht nur positive Entwicklungen nach sich. Soziale Online-Medien bieten auch einen Nährboden für „Inzivilität“, die wir in diesem Beitrag als Kommunikationsform definieren, die interpersonale oder deliberative Normen überschreitet und etwa in Form von aggressiven Nutzerkommentaren, Shitstorms, *Flaming*, *Trolling* oder Hassrede in (teil-)öffentlich zugänglichen Diskussionen beobachtet werden kann (siehe das folgende Kapitel für eine ausführliche Begriffsklärung). Obwohl diese Beobachtung alles andere als neu ist und bereits in frühen Arbeiten zu den sozialpsychologischen Aspekten computervermittelter Kommunikation diskutiert wurde (siehe z.B. Kiesler, Siegel, & McGuire, 1984), scheint das Risiko einer Konfrontation mit inzivilen Diskursen in sozialen Online-Medien zugenommen zu haben (Coe, Kenski, & Rains, 2014; Kaakinen, Oksanen, & Räsänen, 2018; Rost, Stahel, & Frey, 2016; Su u. a., 2018). Eine Analyse von über 243 Millionen Nutzerkommentaren, die unter Beiträgen der Facebook-Seiten von 42 US-Nachrichtenmedien hinterlassen wurden, zeigte jüngst, dass zwischen 25 und 41 Prozent dieser Posts inzivile Kommunikation enthalten (Su u. a., 2018).

Auch für Deutschland lassen sich einige Indikatoren für (eine Zunahme von) Inzivilität finden. So brachte etwa die Einführung des Netzwerkdurchsetzungsgesetzes (NetzDG) – das SNS-Betreiber

verpflichtet, Hassreden, beleidigende, verleumdende und andere rechtswidrige Inhalte innerhalb von 24 Stunden nach Eingang einer Beschwerde zu sperren – eine entsprechende Berichtspflicht der Anbieter mit sich. Die Berichte offenbaren, dass Nutzer*innen von Facebook, Twitter, YouTube und Google+ im Zeitraum von Januar bis Juli 2018 mehr als 500.000 Beiträge gemeldet haben (Gollatz, Riedl, & Pohlmann, 2018). Daneben zeigen Berichte von *jugendschutz.net*, dass die Zahl von Verstößen gegen Jugendschutzbestimmungen im Internet von rund 6.000 registrierten Fällen im Jahr 2015 (M. Glaser, Herzog, Özkilic, & Schindler, 2016) auf über 7.500 im Jahr 2017 angestiegen ist (M. Glaser, Herzog, Özkilic, & Schindler, 2018). Vor diesem Hintergrund ist es auch kaum verwunderlich, dass die überwiegende Mehrheit deutscher Jugendlicher angibt, in sozialen Online-Medien bereits mit extremistischen *Posts* oder Hasskommentaren in Kontakt gekommen zu sein (Reinemann, Nienierza, Fawzi, Riesmeyer, & Neumann, 2019).

Es ist somit keine Übertreibung, wenn man Inzivilität als zentrales Merkmal der Online-Kommunikation beschreibt. Wer in sozialen Online-Medien aktiv ist, trifft mit hoher Wahrscheinlichkeit auf Hass, Beleidigungen und Pöbeleien, die sich sowohl auf einzelne Nutzer*innen und öffentliche Akteure als auch auf ganze gesellschaftliche Gruppen beziehen können. Die aktive Nutzung sozialer Online-Medien macht es jedoch nicht nur wahrscheinlicher, mit Inzivilität in Kontakt zu kommen, sondern auch selbst zum Opfer von entsprechenden Kommunikationsformen zu werden (Costello, Hawdon, & Ratliff, 2017; Ybarra, Mitchell, Wolak, & Finkelhor, 2006). Neben den fraglos schwerwiegenden Auswirkungen auf die jeweiligen Opfer, kann die Konfrontation mit Inzivilität auch (unbeteiligte) Beobachter*innen selbst, sowie deren Wahrnehmungen, Intentionen und eigenes Kommunikationsverhalten beeinflussen. Ein umfassendes Verständnis der Ursachen und Wirkungen von Inzivilität in sozialen Online-Medien ist daher unerlässlich.

Der vorliegende Beitrag widmet sich daher den negativen Aspekten eines Wandels der Sprach- und Debattenkultur in sozialen Online-Medien. Im Rahmen eines narrativen Literaturüberblicks

diskutieren wir sowohl mögliche *Ursachen* der Wahrnehmung einer inzivilen Sprach- und Debattenkultur als auch *Wirkungen* auf die Nutzer*innen sozialer Online-Medien. Der Fokus auf Inzivilität soll dabei jedoch nicht suggerieren, dass die Kommunikation im Internet ausschließlich negative Effekte produziert oder gar provoziert. Er ergibt sich vielmehr aus der Annahme, dass die gesellschaftlichen Auswirkungen in diesem Bereich grundsätzlich als gravierender erscheinen sowie dem Ziel, eine Diskussion über mögliche *Gegenmaßnahmen* anzuregen. Auch solche Möglichkeiten zur Verhinderung bzw. Verminderung von Inzivilität nehmen wir in diesem Beitrag explizit in den Blick und bieten so Ansatzpunkte für die Entwicklung von Strategien zur Verbesserung der diskursiven Qualität.

2.

Begriffsklärung
und Perspektive
des Beitrags

2. Begriffsklärungen und Perspektive des Beitrags

Wie eingangs beschrieben, fokussiert der vorliegende Beitrag auf **Inzivilität** (engl. *incivility*) in sozialen Online-Medien. Bevor wir jedoch auf die Ursachen und Wirkungen von inziviler Kommunikation eingehen, soll zunächst geklärt werden, was genau darunter zu verstehen ist und welche (Sub-)Phänomene wir im Rahmen unserer Analyse in den Blick nehmen.¹

Inzivilität gilt als „notoriously difficult term to define“ (Coe u. a., 2014, S. 660) und wird je nach Schwerpunkt konkreter Studien und Projekte unterschiedlich konzeptualisiert. Unser Fokus liegt zunächst auf Inzivilität, die sich in (teil-)öffentlichen Diskussionen und Debatten auf sozialen Online-Medien manifestiert, somit prinzipiell für andere Nutzer*innen beobachtbar ist und daher ein Problem auf gesellschaftlicher Ebene darstellt bzw. darstellen kann. Trotz dieses Fokus auf öffentliche Kommunikation schließen wir sowohl Inzivilität auf *persönlicher Ebene* („personal-level incivility“; Interaktionen, die sich durch Beleidigungen, Respektlosigkeit und aggressive Kommunikationsformen auszeichnen) als auch Inzivilität auf *öffentlichkeitstheoretischer Ebene* („public-level incivility“; Verstöße gegen kommunikative Normen des bürgerschaftlichen Diskurses sowie demokratische Werte) in unsere Definition mit ein (Chen, 2017; Coe u. a., 2014; Muddiman, 2017; Papacharissi, 2004).

Wir verstehen Inzivilität in sozialen Online-Medien auf dieser Basis als normüberschreitende Kommunikation, die sich auf Normen der interpersonellen und/oder deliberativen Kommunikation bezieht und in (teil-)öffentlich zugänglichen Diskussionen auf sozialen Netzwerkseiten, Foren, Online-Communities, Kommentarsektionen auf Webseiten sowie Videoplattformen oder Wikis beobachtet werden kann. Obwohl der Großteil bisheriger Forschungsarbeiten auf textbasierte Formen von Inzivilität fokussiert (insbesondere

auf schriftliche Kommentare/Posts von Nutzer*innen), ist in der Definition auch der bildbasierte oder auditive Ausdruck von Inzivilität in Form von sogenannten *Memes*, Videoclips oder Soundschnipseln mitgedacht. Im Rahmen dieser übergeordneten Definition lassen sich relevante (Sub-)Phänomene unterscheiden, die zwar alle als inzivil gelten können, sich jedoch insbesondere mit Blick auf die Zielgerichtetheit der Kommunikation, das Ausmaß sowie die zugrundeliegenden Beweggründe kontrastieren lassen.²

Ein insbesondere in den letzten Jahren zunehmend diskutiertes Phänomen ist die **Online-Hassrede** (engl. *hate speech*), die den sprachlichen Ausdruck von Hass gegen gesellschaftliche Gruppen (so zum Beispiel aufgrund einer sexuellen Orientierung, Religionszugehörigkeit, aufgrund des Geschlechts, der Ethnie) beschreibt und sich insbesondere durch die Verwendung von Ausdrücken auszeichnet, die der Verunglimpfung und Herabsetzung eben dieser Gruppen dienen (siehe Blazak, 2009; Meibauer, 2013; Schmitt, 2017; unter dem Begriff schädliche Rede [engl. *harmful speech*] auch Leets & Giles, 1999). Hassrede kann in sozialen Online-Medien verschiedene Formen annehmen, wobei grundsätzlich zwischen offenen, „in your face“ und stärker verdeckten und impliziten Praktiken der Diskriminierung unterschieden werden kann (Ben-David & Matamoros-Fernández, 2016; Borgeson & Valeri, 2004).

Von der Online-Hassrede abzugrenzen sind sogenannte **Shitstorms** (außerhalb des deutschen Sprachraums wird in der Regel von *online firestorms* gesprochen), die durch das plötzliche und „sturmartige“ Auftreten von negativer Kritik an Unternehmen oder im öffentlichen Interesse stehenden Personen charakterisiert sind (siehe z.B. Johnen, Jungblut, & Ziegele, 2018; Pfeffer, Zorbach, & Carley, 2014; Rost u. a., 2016). Inzivilere Kommunikation ist hier jedoch nicht die Folge

von Hass gegen grundlegende Attribute sozialer Akteure (z.B. Ethnie, Geschlecht), sondern vielmehr von öffentlicher Missbilligung moralisch fragwürdiger Handlungen oder Äußerungen bekannter Personen oder Organisationen. Wie der namensgebende Sturm zeichnen sich Shitstorms durch ein plötzliches Auftreten und einen ebenso schnellen Rückgang aus – die Inzivilität entlädt sich somit in einem relativ engen Zeitfenster und ist in der Regel auf ein spezifisches, als unmoralisch empfundenen, Ereignis fokussiert.

Im Gegensatz zur Hassrede gelten *Flaming* und *Trolling* als Formen der Inzivilität, die primär den interaktionalen Normverstoß (persönliche Ebene von Inzivilität) beschreiben und sich in aller Regel nicht auf gesellschaftliche Gruppen, sondern konkrete Einzelpersonen beziehen. Und im Gegensatz zum Begriff des Shitstorms, der stark in einer Makrologik verhaftet ist und vor allem bei einer beobachtbaren Häufung von Inzivilität gegen öffentlich bekannte Akteure/Organisationen genutzt wird, werden die Begriffe *Flaming* und *Trolling* auch dann verwendet, wenn es sich um singuläre und individuell eingesetzte Formen inziviler Kommunikation handelt.

Trolling wird vor allem mit Blick auf den disruptiven Charakter der Kommunikation beschrieben und lässt sich mit der Formel „Stören um des Störens willen“ anschaulich beschreiben (siehe z.B. Buckels, Trapnell, & Paulhus, 2014; Craker & March, 2016; Rieger, Dippolt, & Appel, im Druck). Sogenannte Trolle (d.h. Nutzer*innen, die *Trolling*-Verhalten zeigen) verfolgen das Ziel, andere Nutzer*innen bewusst zu provozieren und so zu emotionalen Reaktionen zu bewegen. Die Auswahl der Opfer erfolgt anders als bei Hassredner*innen nicht systematisch, sondern eher willkürlich und opportunistisch. Im Sinne der Begriffsherkunft – *Trolling* bezeichnet ursprünglich eine spezifische Angeltechnik, bei der ein Köder hinter einem fahrenden Boot hergezogen wird – sind dabei prinzipiell alle Nutzer*innen gefährdet, die bei der Provokation „anbeißen“.

Flaming hingegen wurde zumeist hinsichtlich der Nutzung „aufgeheizter“ – im Sinne von aggressiver und feindseliger – Sprache charakterisiert

(Hmielowski, Hutchens, & Cicchirillo, 2014; Moor, Heuvelman, & Verleur, 2010; O’Sullivan & Flanagan, 2003), ist davon abgesehen aber der am wenigsten klar umrissene Begriff. Insbesondere in frühen Arbeiten war *Flaming* häufig der Oberbegriff für eine Vielzahl inziviler Kommunikationsformen, wird in der wissenschaftlichen Auseinandersetzung jedoch immer seltener genutzt. Nicht zuletzt deshalb argumentieren einige Autor*innen (ausführlich etwa Jane, 2015), dass sich die Forschung weniger darauf fokussieren sollte, mit „chirurgischer Präzision“ (S. 80) nach Definitionen zu suchen, sondern die Aufmerksamkeit vielmehr auf die übergreifenden Probleme richten sollte, die mit Inzivilität in sozialen Online-Medien in Verbindung stehen. Entsprechend wollen auch wir in unserem Beitrag vorrangig auf generalisierbare Ursachen und Wirkungen von Inzivilität eingehen, ohne dabei jedoch die Besonderheiten relevanter (Sub-)Phänomene zu ignorieren.³

- 1 Während ein Wandel der Sprach- und Debattenkultur bzw. Erscheinungsformen von Inzivilität auch aus linguistischer oder sprachphilosophischer Perspektive beleuchtet werden kann, nehmen wir in diesem Beitrag einen kommunikationswissenschaftlichen Standpunkt ein. Angesichts der Interdisziplinarität des Fachbereichs berücksichtigen wir dabei auch Arbeiten aus angrenzenden Bereichen der (Sozial-)Psychologie oder Politikwissenschaft, sofern sie einen explizit sozialwissenschaftlichen Fokus haben und den Gegenstand unserer Analyse teilen.
- 2 An dieser Stelle sei darauf hingewiesen, dass die nachfolgend beschriebenen (Sub-)Phänomene – ähnlich wie der von uns als Oberbegriff gewählte Term der Inzivilität – in der wissenschaftlichen Auseinandersetzung sehr unterschiedlich definiert und gemessen werden. Unsere Definitionen beschreiben daher gewissermaßen den „kleinsten gemeinsamen Nenner“.
- 3 Explizit ausklammern werden wir in diesem Beitrag das Phänomen des Cybermobbings bzw. Cyberbullyings (für einen Überblick siehe Fawzi, 2015), da dieses in aller Regel auf eine sehr spezifische Form der Inzivilität unter Kindern und Jugendlichen beschränkt ist und Wirkungen primär aus einer Opfer-Perspektive diskutiert.

3.

Ursachen für die
Wahrnehmung
von mehr
Inzivilität
in sozialen
Online-Medien

3. Ursachen für die Wahrnehmung von mehr Inzivilität in sozialen Online-Medien

Während die kommunikationswissenschaftliche Forschung keine Aussagen darüber treffen kann, *wie* oder *warum* sich Sprache im linguistischen Sinne verändert, kann sie dennoch Antworten darauf finden, warum von vielen Beobachter*innen in sozialen Online-Medien eine negative Veränderung der Sprach- und Debattenkultur sowie ein Anstieg von Inzivilität wahrgenommen wird.

Dabei wäre zunächst die **(1) gesteigerte Sichtbarkeit und öffentliche Zugänglichkeit von Debatten und Diskursen** zu nennen (siehe z.B. Blom, Carpenter, Bowe, & Lange, 2014, S. 1316–1317; Chen, 2017, S. 58; Coe u. a., 2014, S. 658; Neubaum & Krämer, 2017, S. 464; Neuberger, 2017, S. 104). Insbesondere auf sozialen Netzwerkseiten haben die Nutzer*innen die Möglichkeit, eine Vielzahl von Unterhaltungen zu beobachten: Seien es Diskussionen unter den Tweets von Donald Trump, Debatten in den Kommentarspalten von auf Facebook geteilten Nachrichtenartikeln oder Threads zu aktuellen politischen Themen in Online-Communities wie Reddit. Die „socially mediated publicness“ (Baym & boyd, 2012) von SNS ermöglicht Nutzer*innen dabei jedoch nicht nur das Zuschauen, sondern auch die aktive eigene Teilnahme. Wer über ein entsprechendes Benutzerkonto verfügt, kann sich nahezu aufwandslos selbst in Diskussionen einschalten. Ein Interesse an der Teilnahme an oder zumindest der Beobachtung von Diskursen in sozialen Online-Medien scheint vor allem bei akuten gesellschaftlichen Veränderungen oder disruptiven Ereignissen (z.B. nach Terroranschlägen) zuzunehmen. Studien in diesem Bereich suggerieren, dass derartige Situationen oder Vorfälle zudem zu einem Anstieg von Inzivilität führen können (Kaakinen u. a., 2018; Oksanen u. a., 2018): So ist etwa die Verwendung von Hassrede für manche Nutzer*innen ein (Aus-)Weg, um mit der empfundenen

Unsicherheit und Ungerechtigkeit umzugehen. Insbesondere Mitglieder von – oft grob generalisierten – Fremdgruppen müssen in solchen Situationen als Sündenbock herhalten und werden daher mit Inzivilität konfrontiert. Nach islamistisch motivierten Terroraktivitäten werden beispielsweise Muslim*innen häufig zu Opfern von Hassrede, da sie aufgrund ihrer religiösen Zugehörigkeit mit den Attentätern assoziiert oder gleichgesetzt werden (siehe z.B. Awan, 2014, 2016; Chetty & Alathur, 2018).

Eng mit diesem ersten Punkt verbunden sind **(2) vereinfachte Möglichkeiten zur schnellen und weiten Verbreitung von Inhalten**. Die Affordanzen sozialer Online-Medien regen in besonderem Maße dazu an, Informationen und Diskussionen weiterzubreiten und mit anderen zu teilen (Kümpel, Karnowski, & Keyling, 2015; Osatuyi, 2013), was die Sichtbarkeit – insbesondere von hitzigen Debatten – weiter erhöht. So spricht eine Vielzahl an Studien dafür, dass vor allem emotional aufgeladene und erregende Inhalte in sozialen Online-Medien eine weite Verbreitung erfahren (z.B. Berger, 2011; Hasell & Weeks, 2016; Stieglitz & Dang-Xuan, 2013). Im Kontext der US-Präsidentenwahl im Jahr 2012 etwa finden Hasell und Weeks (2016), dass Ärger positiv mit dem Teilen von politischen Informationen in Verbindung steht: Nutzer*innen, die aufgrund ihrer Mediennutzung wütend auf den jeweils gegnerischen Präsidentschaftskandidaten waren, teilten und kommentierten häufiger Beiträge in sozialen Online-Medien – was die empfundene Wut somit auch für (viele) weitere Nutzer*innen sichtbar machte. Komplementiert werden diese früheren Befunde durch eine aktuelle Analyse der auf Facebook meistgeteilten und -kommentierten Nachrichtenbeiträge. Weitverbreitet und diskutiert wurden im ersten Quartal 2019 vorrangig jene

Meldungen, die Angst induzieren oder polarisieren, beispielsweise *Posts* über Abtreibungen oder illegale Immigration (Owen, 2019). Dies wiederum lässt den dazugehörigen Diskussionen mutmaßlich noch mehr Aufmerksamkeit zuteilwerden, da die Algorithmen sozialer Netzwerke jene Beiträge, die viel Auseinandersetzung (engl. *engagement*) in Form von *Shares*, *Likes* oder Kommentaren provozieren, prominenter im Neuigkeiten-Bereich anderer Nutzer*innen platzieren (DeVito, 2017; Owen, 2018). Argumentiert wird vor diesem Hintergrund auch, dass soziale Online-Medien Verbreiter*innen von Hass unabsichtlich dazu verhelfen können, ihre aggressiven Ansichten und Aussagen in Diskussionen einzuschleusen. Klein (2012) spricht in diesem Zusammenhang von „Information Laundering“ und beschreibt damit die Idee, dass soziale Online-Medien Hass-Gruppen und vergleichbare Akteure dazu befähigen, ihre „illegale Währung“ (in diesem Fall: Hassbotschaften) reinzuwaschen und in Umlauf zu bringen. Am Beispiel von extremen politischen Parteien und Organisationen in Spanien zeigen Ben-David und Matamoros-Fernández (2016) diesbezüglich, wie sowohl offene Hassrede und eher subtilere Formen von Diskriminierung über Facebook in Zirkulation gebracht werden. Die Autor*innen diskutieren dabei auch die Rolle der SNS-Anbieter kritisch, die einerseits die Verbreitung von Hassrede und Inzivilität offiziell verbieten, andererseits aber durch *Share*-, *Like*- und Kommentar-Features die ideale Infrastruktur für „the circulation and accumulation of subtle associations of hate and discrimination“ (ebd., S. 1187) bereitstellen.

Sprechen diese ersten beiden Aspekte vorrangig für eine stärkere *Wahrnehmung* eines Wandels der Sprach- und Debattenkultur hin zu mehr Inzivilität, gibt es Hinweise darauf, dass **(3) veränderte Kommunikationsbedingungen und (Gruppen-) Normen** in sozialen Online-Medien negatives Kommunikationsverhalten in der Tat forcieren können. Viel diskutiert wurde in diesem Kontext die Rolle von Anonymität (siehe z.B. Brown, 2018; Matzner, 2016; Puryear & Vandello, 2018; Rösner & Krämer, 2016; Rowe, 2015a; Santana, 2014; Zimmerman & Ybarra, 2016). Auch wenn einige soziale Online-Medien die Verwendung von

Klarnamen erbitten (und sich viele Nutzer*innen – etwa bei Facebook – angesichts des Vernetzungsgedankens daran orientieren), gibt es vielfältige Möglichkeiten, die eigene Identität zu verbergen. Schon früh wurde argumentiert, dass die (relative) Anonymität computervermittelter Kommunikationsräume zu Enthemmung und einer Verringerung des Verantwortungsbewusstseins beitrage und Nutzer*innen online Handlungen ausführen, die sie in der persönlichen Kommunikation nie erwägen würden (Postmes, Spears, & Lea, 1998; Suler, 2004). Der Zusammenhang von Anonymität und unmoralischem Verhalten wurde darüber hinaus jedoch auch mit Prozessen der Deindividuation erklärt (Kiesler u. a., 1984; Reicher, Spears, & Postmes, 1995; Spears & Lea, 1994). So geht das in der Sozialpsychologie entwickelte SIDE-Modell (*social identity model of deindividuation effects*) davon aus, dass Anonymität die Zugehörigkeit zu sozialen Gruppen bewusster machen kann und so die Orientierung an (wahrgenommenen) Gruppennormen steigert. Ein abschätziger Kommunikationsstil erscheint in manchen Foren oder Kommentarspalten als „ganz normal“ und prägt so auch das Verhalten von neuen Diskussionsteilnehmer*innen (Chen & Lu, 2017; Hsueh, Yogeewaran, & Malinen, 2015; Sood, Antin, & Churchill, 2012; Stroud, Scacco, Muddiman, & Curry, 2015; Sukumaran, Vezich, McHugh, & Nass, 2011). Analog zur sogenannten Broken-Windows-Theorie (Wilson & Kelling, 1982), nach der ein Zusammenhang zwischen dem Verfall von Stadtteilen und der dort beobachteten Delinquenz besteht, kann so auch für inzivile Online-Diskussionen ein negativer Ausstrahlungseffekt auf andere beteiligte Akteure angenommen werden. Wird indes eine positive und zivile Diskussionsatmosphäre geschaffen, so verbessert sich auch das Kommunikationsverhalten der Teilnehmer*innen sichtlich (Fredheim, Moore, & Naughton, 2015). Ob Anonymität per se tatsächlich mit mehr Inzivilität in der Kommunikation in Verbindung steht, ist nicht eindeutig zu beantworten. Während einige Studien für einen solchen Zusammenhang sprechen (Rowe, 2015b; Santana, 2014; Zimmerman & Ybarra, 2016), finden andere keine Effekte von Anonymität auf die Diskussionsqualität oder die Nutzung aggressiver Sprache (Berg, 2016; Rösner & Krämer, 2016).

Die in sozialen Online-Medien erlebbare Anonymität geht jedoch mit weiteren Besonderheiten der Kommunikationssituation einher, die ebenfalls (negative) Auswirkungen auf Struktur und Inhalt von Debatten haben können. Da die meisten Diskussionen auch in Online-Settings textbasiert stattfinden, fehlt es an zentralen sozialen Hinweisreizen, die im *Face-to-Face*-Kontext der Regulation von Kommunikation dienen (siehe z.B. Brown, 2018; Dutton, 1996; Joinson, 2007; Kiesler u. a., 1984; Lapidot-Lefler & Barak, 2012; Postmes u. a., 1998; Suler, 2004). Die Unsichtbarkeit der Gesprächspartner*innen, das Fehlen von Augenkontakt sowie der Wegfall von Gestik, Mimik und Intonation können dazu beitragen, dass Inzivilität von den Nutzer*innen sozialer Online-Medien als weniger schwerwiegend und verletzend wahrgenommen wird. Entsprechend gibt es deutlich weniger Hemmnisse, sich aggressiver, beleidigender oder hasserfüllter Kommunikationsformen zu bedienen.

Fokussiert man auf die Ursachen eines Wandels der Sprach- und Debattenkultur, dürfen insbesondere aus kommunikationswissenschaftlicher Perspektive schließlich auch **(4) motivationale Aspekte und Persönlichkeitsmerkmale** nicht außer Acht gelassen werden. Ein Wissen darüber, *warum* und *welche* Nutzer*innen inzivile und aggressive Kommentare schreiben, Hassbotschaften verbreiten, oder trollen, ist daher ebenso wichtig wie der Blick auf die veränderten Informationsumgebungen und Kommunikationsbedingungen. Ein Grund dafür, warum Diskussionen in sozialen Online-Medien schnell hitzig oder aggressiv werden können, lässt sich bereits aus generellen Motiven des Kommentierens ableiten: Viele Nutzer*innen verfassen Kommentare, wenn sie Kritik oder Widerspruch üben oder sich von aufgestauten Emotionen befreien wollen – Wut, Ärger und Empörung spielen dabei eine bedeutsame Rolle und spiegeln sich häufig auch im Kommunikationsverhalten wider (Moor u. a., 2010; Springer, 2014; Upadhyay, 2010; Ziegele, 2016). Besonders wahrscheinlich ist ein solches aggressives Kommunikationsverhalten dann, wenn in Diskussionen ein Angriff auf die eigenen (politischen) Ansichten wahrgenommen wird (Hutchens, Cicchirillo, & Hmielowski, 2015)

oder die eigene soziale Identität als bedroht erlebt werden (J. Glaser, Dixit, & Green, 2002). Während Inzivilität in diesen Fällen eher als unbeabsichtigte Folge eines konkreten kommunikativen Ziels entsteht, gibt es auch Hinweise darauf, dass inzivile Kommunikationsformen bewusst eingesetzt werden, um sich von einer – wie auch immer im Einzelfall definierten – Fremdgruppe abzusetzen und so wiederum den Zusammenhalt in der Eigengruppe zu stärken (Upadhyay, 2010). Dabei spielt auch eine Rolle, wie stark die Eigengruppe symbolisch bedroht wird: In rassistischen Chatträumen in den USA etwa wurden Themen wie die interkulturelle („gemischtrassige“) Ehe oder der Zuzug von Afroamerikaner*innen in „weiße“ Nachbarschaften als Auslöser von Inzivilität identifiziert (J. Glaser u. a., 2002). Motivationen in Zusammenhang mit der Abgrenzung „Wir“ gegen „die Anderen“ zeigen sich auch für die Verbreitung von Online-Hassreden: Die verbale Herabwürdigung einer Fremdgruppe zielt explizit auf deren Ausgrenzung, bestätigt aber gleichermaßen auch die eigene Gruppenidentität (McNamee, Peterson, & Peña, 2010; Schmitt, 2017). Interessanterweise kann aber auch das couragierte Einschreiten in hasserfüllte Online-Diskussionen (mehr) Hass produzieren: Wer sich an Diskussionen mit Hassrede-Anteilen beteiligt, wird mit größerer Wahrscheinlichkeit selbst zum Opfer (Costello u. a., 2017).

Mit Blick auf Persönlichkeitsmerkmale und situative Einflüsse zeigt sich zudem, dass eine Disposition für verbale Aggression (Cicchirillo, Hmielowski, & Hutchens, 2015; Hmielowski u. a., 2014), schlechte Stimmung (Cheng, Bernstein, Danescu-Niculescu-Mizil, & Leskovec, 2017), die Suche nach Aufmerksamkeit (Shachaf & Hara, 2010) sowie Psychopathie (im Sinne eines mangelnden Empathievermögens) und Sadismus (Buckels u. a., 2014; Craker & March, 2016) mit *Trolling*-Verhalten und Inzivilität in Verbindung stehen. Einige Nutzer*innen scheinen darüber hinaus durch Unterhaltungsmotive angetrieben zu sein: So werden in der Literatur auch Spaß und Nervenkitzel sowie die Freude am Schädigen anderer als Auslöser für *Trolling* und Online-Hassrede diskutiert (Schmitt, 2017; Shachaf & Hara, 2010). Erjavec und Kovačič (2012) beschreiben

diese Nutzer*innen auch als „die Spieler“ (engl. *The Players*), die nicht auf Basis innerer Überzeugungen handeln, sondern lediglich Unterhaltung aus der Demütigung anderer ziehen.

Zusammenfassend lässt sich mithin festhalten, dass eine Kombination aus individuellen Motivlagen, den für soziale Online-Medien spezifischen Kommunikationsbedingungen sowie der gesteigerten Sichtbarkeit und den einfachen Möglichkeiten zur Weiterverbreitung von Debatten und Diskursen dazu beigetragen hat, dass wir in Kommunikationsräumen im Internet verstärkt Negativität und Inzivilität wahrnehmen können. Doch was bedeutet das für diejenigen, die diesen Diskussionen ausgesetzt sind? Welche Wirkungen hat die Beobachtung von Hass und Respektlosigkeit – sowohl auf (unbeteiligte) Nutzer*innen als auch die von Inzivilität Betroffenen? Diesen Fragen wird nun im nächsten Teil des Beitrags nachgegangen.

4. Wirkungen von Inzivilität in sozialen Online-Medien

4. Wirkungen von Inzivilität in sozialen Online-Medien

Beschäftigt man sich in der Kommunikationswissenschaft mit der Wirkung von „klassischen“ Medieninhalten (z.B. journalistische Texte, Videobeiträge) oder anderen, zum Teil unmittelbar damit in Verbindung stehenden Kommunikaten (z.B. Nutzerkommentare auf Facebook, Twitter-Tweets), lassen sich grundsätzlich drei Wirkungsbereiche unterscheiden (siehe Schenk, 2007, S. 767): Wirkungen auf (1) Einstellungen, Meinungen und Verhalten, (2) Stimmungen und Emotionen sowie (3) Vorstellungen und Wissen. Auch Inzivilität in Diskussionen und Debatten auf sozialen Online-Medien kann derartige Wirkungen bei Teilnehmer*innen und Beobachter*innen hervorrufen und beispielsweise die Wahrnehmung von Themen, die emotionale Erregung oder Vorstellungen von gesellschaftlichen Gruppen beeinflussen (ausführlich siehe unten). Die obige Dreiteilung soll jedoch nicht suggerieren, dass sich Effekte immer nur in einem Teilbereich zeigen – Medienwirkungen sind in aller Regel komplex, bedingen und beeinflussen sich gegenseitig. Zudem trifft jedes Kommunikat stets auf Individuen mit je unterschiedlichen Persönlichkeitsmerkmalen und Voreinstellungen, die sich in je unterschiedlichen Situationen und Stimmungen befinden.

Damit Medieninhalte und andere Kommunikate auf Nutzer*innen wirken können, müssen gewisse Voraussetzungen erfüllt sein. Die schiere Menge an uns umgebenden (medialen) Reizen erfordert es, eine Auswahl zu treffen: Nur Inhalte, die willkürlich oder unwillkürlich die Aufmerksamkeit einer Nutzer*in erregen, sind überhaupt in der Lage, Wirkungen zu entfalten (Schweiger, 2007, S. 137–138). Aufmerksamkeitsprozesse sind das Resultat eines individuell verschiedenen Wechselspiels aus Reiz/Inhalt und Reaktion/Person und somit nicht für alle Nutzer*innen gleich. Kurzum: Was einigen sofort ins Auge springt, nehmen andere vielleicht erst auf den zweiten oder dritten Blick wahr. Trotz aller subjektiven Unterschiede

gibt es jedoch bestimmte Reize, die „kollektive Relevanz“ besitzen und bei einer Vielzahl von Nutzer*innen Aufmerksamkeit generieren können (Eilders, 1997, S. 92 ff.). Zurückführen lässt sich dies unter anderem auf evolutionäre Aspekte (z.B. indiziert das Merkmal „Schaden“ in Kommunikaten ein Gefahrenpotenzial, welches es zu untersuchen gilt), geteilte Werte und Normen sowie stärker automatisierte Verarbeitungsprozesse (ebd.). Inzivilität im von uns definierten Sinne verfügt potenziell gleich über mehrere Merkmale, die intersubjektiv die Beurteilung von Relevanz prägen können. So sind etwa Online-Hassrede oder Shitstorms häufig mit etwaigem oder tatsächlichem Schaden verbunden und zeichnen sich nicht zuletzt auf diskursiver Ebene durch Kontroverse bzw. Konflikthaftigkeit aus. Befunde zur Wahrnehmung von Nachrichtenbeiträgen zeigen, dass diese Faktoren (bei Vorkommen in journalistischen Texten) Relevanzurteile der Rezipient*innen positiv beeinflussen: Texte, die kontrovers sind und Schäden bzw. negative Konsequenzen thematisieren, werden als wichtiger und bedeutsamer wahrgenommen (Weber & Wirth, 2013). Demnach ist anzunehmen, dass auch in sozialen Online-Medien beobachtete inzivile Diskussionen eine gewisse kollektive Relevanz besitzen, entsprechend aufmerksam(er) wahrgenommen werden und so weiterführende Wirkungen auf die Beobachter*innen haben können. Das genaue Ausmaß ist dabei jedoch von Person zu Person verschieden.

Subjektivität kennzeichnet nicht nur die Aufmerksamkeit gegenüber Inzivilität, sondern auch die *inhaltliche* Wahrnehmung des Kommunikats. Wie oben bereits beschrieben, ist (In-)Zivilität in der Literatur ganz unterschiedlich definiert worden, was auch daran liegt, dass es „very much in the eye of the beholder“ (Herbst, 2010, S. 3) liegt, ob eine Debatte oder Diskussion als inzivil oder aggressiv wahrgenommen wird (siehe auch Coe u. a., 2014, S. 660). Ähnlich wie die beschrie-

benen kollektiven Relevanzzuweisungen gibt es aber auch bei der Wahrnehmung von Inzivilität übergreifende Tendenzen, die sich bei einer Mehrheit von Beobachter*innen zeigen. Mit einem spezifischen Fokus auf *politische* Inzivilität (d.h., Inzivilität in politischen Diskussionen zwischen Politiker*innen, den Medien und Bürger*innen) haben Stryker und Kolleg*innen (2016) untersucht, inwieweit Konsens darüber besteht, welche Kommunikationsformen als inzivil wahrgenommen werden. Dafür haben sie die Teilnehmer*innen ihrer Studie mit 23 potenziell inzivilen Handlungen konfrontiert (u.a. Nutzung vulgärer Sprache, Beschimpfungen, Androhung von Gewalt, persönliche Angriffe) und sie gebeten, die Inzivilität dieser Handlungen zu bewerten. Dabei zeigte sich zunächst, dass die von den Forscher*innen vorgelegten Handlungen von über 75 Prozent der Teilnehmer*innen als zumindest etwas inzivil wahrgenommen wurden. Die größte Einigkeit zeigte sich bei der Androhung von Gewalt sowie der Nutzung von rassistischen, religiösen, ethnischen oder sexistischen Beleidigungen – diese Handlungen wurden von 82 bis 87 Prozent der Befragten als sehr inzivil empfunden.¹ Im Gegensatz dazu wurden Attacken auf die inhaltlichen Standpunkte eines politischen Gegners von der Mehrheit nicht als inzivil wahrgenommen, was auf die Unterschiede zwischen themen- und personenbezogenen Angriffen verweist. Obschon die Studie erste Hinweise auf die (geteilte) Wahrnehmung von Inzivilität zulässt, ist sie durch den ausschließlichen Fokus auf politische Inzivilität sowie den fehlenden expliziten Bezügen zu Merkmalen internetbasierter Kommunikation nicht vollständig auf die Situation in sozialen Online-Medien übertragbar.

Diese Lücke schließen Kenski und Kolleg*innen (2017), die einerseits die Frage aufwerfen, ob und inwiefern fünf in Online-Nutzerkommentaren vorkommende Formen von Inzivilität (Beschimpfungen, Vulgarität, Lügenvorwürfe, Verunglimpfung von Sprache, Verleumdung) als unterschiedlich (in-)zivil wahrgenommen werden, andererseits aber auch untersuchen, ob soziodemographische Charakteristika und Persönlichkeitsmerkmale diese Wahrnehmungen beeinflussen. Im Ansatz ähnlich zu dem Vorgehen von Stryker und Kol-

leg*innen (2016) haben die Forscher*innen den Befragten verschiedene Statements vorgelegt, die als echte Kommentare auf einer Nachrichten-Webseite ausgegeben wurden. Diese Statements sollten dann in Hinblick auf ihre Inzivilität bewertet werden. Beleidigungen wurden auch in dieser Studie als inzivilste Form der Kommunikation wahrgenommen, dicht gefolgt von vulgären Äußerungen. Die Befunde offenbaren darüber hinaus jedoch auch, welche Eigenschaften einer Person mit dem Empfinden von Inzivilität in Zusammenhang stehen. Diesbezüglich zeigte sich, dass Frauen die vorgelegten Statements grundsätzlich als inzivil beurteilen als Männer. Gleichmaßen scheinen Personen, bei denen das Persönlichkeitsmerkmal Verträglichkeit stärker ausgeprägt ist (d.h., die sich als kooperativ und mitfühlend begreifen), mehr Inzivilität in den Kommentaren wahrzunehmen. Es lässt sich also schlussfolgern, dass – obwohl bestimmte inzivile Kommunikationsformen von fast allen Nutzer*innen als solche wahrgenommen werden – Merkmale wie Geschlecht und spezifische Persönlichkeitseigenschaften die Beurteilung von Inzivilität beeinflussen können. Dies wiederum kann in einem nächsten Schritt auch weiterführende Wirkungen der (inzivilen) Aussagen verändern.

Die Wirkung von Inzivilität wurde bislang vor allem im Kontext von Nutzerkommentaren auf Nachrichten-Webseiten untersucht. Ausgangspunkt vieler Studien war häufig die Frage, ob und inwiefern inzivile Äußerungen in Kommentaren für die journalistische Arbeit relevante Wahrnehmungen und Verhaltensintentionen der Rezipient*innen beeinflussen. Empirische Untersuchungen haben wiederholt gezeigt, dass Nutzerkommentare – zunächst ganz unabhängig vom Grad ihrer (In-)Zivilität – individuelle Vorstellungen, Meinungen und Verhaltensweisen beeinflussen können (für einen Überblick siehe Ksiazek & Springer, 2018; Springer & Kümpel, 2018). In sozialen Online-Medien beobachtbare Meinungsäußerungen scheinen also in der Tat wichtige Anhaltspunkte dafür zu liefern, wie (kontroverse) Themen oder die Berichterstattung selbst beurteilt werden können. Inzivilität in Kommentaren wurde vor allem deshalb häufig untersucht, da – wie eingangs beschrieben – von einem hohen Anteil inzivilen

Kommunikationsformen ausgegangen werden muss und negative Wirkungen hier als besonders wahrscheinlich erscheinen.

Die verfügbaren Studien zu den Effekten inziviler Nutzerkommentare sozialer Online-Medien lassen sich primär drei Bereichen zuordnen. So wurden hauptsächlich Einflüsse auf (1) medien- bzw. journalismusbezogene Perzeptionen (z.B. Wahrnehmung journalistischer Qualität, Wahrnehmung von in der Berichterstattung verhandelten Themen, Wahrnehmung von Einseitigkeit und Verzerrungen), (2) Emotionen und (aggressive) Verhaltensintentionen sowie (3) das Kommunikationsverhalten der Nutzer*innen adressiert. Daneben gibt es weitere Untersuchungen, die sich mit der Wirkung auf die von Inzivilität Betroffenen beschäftigen und in den Blick nehmen, was die Konfrontation mit inziviler Kommunikation bei den eigentlichen Opfern bewirkt.

Mit Blick auf den ersten Teilbereich, **(1) medien- bzw. journalismusbezogene Perzeptionen**, zeigt sich zunächst, dass inzivile Kommentare beeinflussen können, wie Nutzer*innen die allgemeine Qualität, Objektivität oder Glaubwürdigkeit von Nachrichtenbeiträgen wahrnehmen. Eine Studie von Prochazka, Weber und Schweiger (2018) offenbart diesbezüglich, dass ein und derselbe Nachrichtenartikel als qualitativ minderwertiger wahrgenommen wird, wenn er von inzivilen Kommentaren begleitet wird (ähnlich für Kommentare mit negativer Valenz auch Dohle, 2018; Kümpel & Springer, 2016; Kümpel & Unkel, im Druck). Interessanterweise führt aber die Präsenz ziviler Kommentare umgekehrt nicht zu einer Verbesserung von Qualitätswahrnehmungen: Tatsächlich scheint bereits die schiere Präsenz von Kommentaren zu bewirken, dass Nutzer*innen die Qualität journalistischer Beiträge negativer evaluieren (Prochazka u. a., 2018). Zu ähnlich ernüchternden Befunden kommen auch Houston, Hansen und Nisbett (2011) sowie Anderson und Kolleg*innen (2018), die im Kontext von Nutzerkommentaren auf Nachrichten-Webseiten bzw. Blogs dokumentieren, dass Inzivilität in Kommentaren die Berichterstattung über politische Akteure sowie objektiv neutrale Blogbeiträge als verzerrter erscheinen lässt. Waddell (2018) identifiziert darüber hinaus

negative Auswirkungen auf die Wahrnehmung der Glaubwürdigkeit des kommentierten Artikels sowie auf die Beurteilung der Wichtigkeit bzw. Relevanz des im Artikel behandelten Themas. Des Weiteren gibt es Hinweise darauf, dass inzivile Kommentare zur Polarisierung von Wahrnehmungen und Meinungen beitragen können (Anderson, Brossard, Scheufele, Xenos, & Ladwig, 2014; Suhay, Bello-Pardo, & Maurer, 2018). Am Beispiel der Berichterstattung zu Nanotechnologie zeigen etwa Anderson und Kolleg*innen (2014), dass bei Individuen, die Nanotechnologie kritisch gegenüberstehen, die Konfrontation mit inzivilen Kommentaren dazu führt, dass die Technologie als (noch) risikoreicher wahrgenommen wird. Die Befunde lassen somit insgesamt keine positiven Schlussfolgerungen für Journalist*innen und Medienanbieter zu. Die experimentelle Forschung zu Inzivilität und Negativität in Nutzerkommentaren zeigt einhellig, dass die Wahrnehmung redaktioneller Angebote selbst dann durch Meinungsäußerungen in sozialen Online-Medien beeinflusst wird, wenn die Inhalte keine begründete Beanstandung zulassen. Dies wiederum könnte langfristig in einen Vertrauensverlust² sowie einen Anstieg von Medienverdrossenheit münden und so den demokratischen Diskurs (noch) weiter gefährden: „A lack of trust in government, unwillingness to associate with people who do not share one’s views, and increased skepticism of the media have the potential to do damage to the way our democracy functions“ (Han, Brazeal, & Pennington, 2018, S. 3).

Neben den Auswirkungen auf medien- bzw. journalismusbezogene Perzeptionen wurden in weiteren Forschungsarbeiten mögliche Effekte von Inzivilität auf **(2) Emotionen und (aggressive) Verhaltensintentionen** untersucht. Die zentrale Frage in diesem zweiten Teilbereich ist also, ob die Konfrontation mit inziviler Kommunikation zu Ärger, negativen Emotionen oder eigenem aggressiven Verhalten beitragen kann. Tatsächlich zeigen sich auch hier die schon im ersten Teilbereich festgestellten Ausstrahlungseffekte. Wer in sozialen Online-Medien inzivilen Kommentaren ausgesetzt ist, berichtet über ein Mehr an negativen Emotionen und Wutgefühlen (Chen & Lu, 2017; Gervais, 2015; Wang & Silva, 2018). Gervais

(2015) findet zudem heraus, dass dies insbesondere dann der Fall ist, wenn Nutzer*innen mit inzivilen Beiträgen konfrontiert werden, die sich gegen ihr eigenes politisches Lager richten: Wer das Gefühl hat, dass die eigene Seite angegriffen wird, ist nach der Rezeption inziviler Kommentare gekränkt und verärgert. Eine Studie von Rösner, Winter und Krämer (2016) deutet des Weiteren darauf hin, dass die Rezeption aggressiver Nutzerkommentare auch bei unbeteiligten Beobachter*innen zu aggressiven Gedanken führen kann, während Chen und Lu (2017) gar einen Zusammenhang mit aggressiven Verhaltensintentionen feststellen. Online-Inzivilität scheint somit nicht nur Gemütslage und Gedanken zu beeinflussen, sondern potenziell auch behaviorale Konsequenzen nach sich zu ziehen.

Eine spezifische Form von behavioralen Konsequenzen nimmt der dritte der oben identifizierten Teilbereiche in den Blick: die Auswirkungen von inziviler Kommunikation auf **(3) das Kommunikationsverhalten der Nutzer*innen**. Dieser Aspekt wurde in diesem Bericht bereits bei den *Ursachen* des Wandels der Sprach- und Debattenkultur angesprochen, kann aber auch als Wirkung von inziviler Kommunikation konzeptualisiert und untersucht werden. Die im vorhergehenden Absatz beschriebene Studie von Chen und Lu (2017) etwa hat nicht nur gezeigt, dass inzivile Kommentare zu mehr negativen Emotionen und aggressiven Intentionen führen, sondern auch, dass Inzivilität in Kommentaren es wahrscheinlicher macht, dass die Nutzer*innen in eigenen kommunikativen Beiträgen ebenfalls auf Inzivilität zurückgreifen. Ähnlich finden Hsueh und Kolleg*innen (2015), dass in Diskussionsumgebungen mit vorurteilsbehafteten Kommentaren (im konkreten Fall ging es um Ressentiments gegenüber Asiat*innen) auch neu hinzukommende Teilnehmer*innen Kommentare verfassen, die solche Vorurteile beinhalten. Allerdings scheint im Bereich Kommunikationsverhalten auch ein umgekehrter (positiver) Effekt zu bestehen: Beobachten Nutzer*innen zivile Diskussionen, so gleichen sie ihre eigenen Äußerungen an diese produktiven Gesprächsnormen an und verhalten sich gleichermaßen zivil (siehe z.B. Han & Brazeal, 2015; Sukumaran u. a., 2011). Die Art

der Kommunikation in sozialen Online-Medien kann also auf eine Art des Beobachtungslernen zurückgeführt werden (ebd.): Je nach prävalenter Diskussionsatmosphäre kann demnach mit negativen *oder* positiven Effekten gerechnet werden. Dies wiederum verdeutlicht die Wichtigkeit eines aktiven *Community Managements* und Moderationsstrategien, denen wir uns bei der Diskussion von Gegenmaßnahmen an späterer Stelle noch ausführlicher widmen werden.

Während die genannten Bereiche allesamt auf die Auswirkungen auf *ordinary citizens* fokussieren, widmen sich einige Studien auch der Frage, wie sich die Konfrontation mit Inzivilität auf spezifische Personengruppen auswirkt. Besonderes Interesse galt dabei den Einflüssen auf Journalist*innen, da diese im Kontext von Kommentarsektionen auf Nachrichten-Webseiten oder SNS-Profilen häufig das primäre Ziel von inzivilen Kommunikationsformen sind (Binns, 2017; Chen u. a., 2018; Nilsson & Örnebring, 2016; Obermaier, Hofbauer, & Reinemann, 2018; Post & Kepplinger, 2019; Preuß, Tetzlaff, & Zick, 2017). Eine Studie unter schwedischen Journalist*innen zeigte bereits im Jahr 2013, dass drei Viertel der Befragten bereits beleidigende oder missbräuchliche Kommentare erhalten haben (Nilsson & Örnebring, 2016). Ganz ähnliche Verhältnisse wurden für Deutschland beobachtet: 74 Prozent der von Post und Kepplinger (2019) befragten Journalist*innen mussten bereits eigene Erfahrungen mit inzivilen Kommentaren sammeln. Auch schwerwiegende Varianten in Form von Hassrede scheinen mittlerweile zum journalistischen Alltag zu gehören und werden von Journalist*innen als wachsendes Problem wahrgenommen (Obermaier u. a., 2018). Auf individueller Ebene kann die Konfrontation mit Inzivilität dazu führen, dass sich Journalist*innen eingeschüchtert fühlen und versuchen, die Berichterstattung über bestimmte konfliktbeladene Themen zu vermeiden oder zumindest weniger konfliktreich darzustellen (Binns, 2017; siehe auch European Federation of Journalists, 2017; Post & Kepplinger, 2019). Auch fühlen sie sich häufiger angegriffen, zweifeln eher an ihrem Zielpublikum und erleben mehr Ärger und negative Emotionen (Obermaier u. a., 2018; Post & Kepplinger, 2019). Um mit Inzivilität

umzugehen, setzen betroffene Journalist*innen vor allem emotionale Bewältigungsstrategien ein und suchen das Gespräch mit Kolleg*innen, Freund*innen oder der eigenen Familie (Obermaier u. a., 2018; Preuß u. a., 2017). Interessanterweise scheinen einige Journalist*innen inzivile und hasserfüllte Reaktionen auf ihre Beiträge aber auch als „professional success“ (Post & Kepplinger, 2019) zu interpretieren und positive Gefühle daraus zu ziehen: Der Angriff von außen scheint innerlich zu bestätigen, dass die journalistische Arbeit richtig und wichtig ist (ebd.; Obermaier u. a., 2018).

Es gibt deutliche Hinweise darauf, dass vor allem weibliche Medienschaffende zu den Opfern von inziviler Online-Kommunikation zählen (Adams, 2018; Chen u. a., 2018). Dies wiederum reflektiert generelle Tendenzen in sozialen Online-Medien. Wie Sobieraj (2018) feststellt, werden nicht nur Journalistinnen, Wissenschaftlerinnen oder Politikerinnen online zu Zielen von Hass, Beleidigungen und verbalen Misshandlungen. Auch „ganz normale“ Nutzerinnen von SNS, Online-Games oder Internet-Communities „find themselves on the receiving end of vitriolic, gender-based backlash“ (S. 1701; siehe auch Fox & Tang, 2014; KhosraviNik & Esposito, 2018). Die Auswirkungen dieser Online-Misogynie reichen von individuellen Effekten der Selbstzensur, Einschüchterung und Unsicherheit hin zum vollständigen Rückzug aus öffentlichen Online-Diskursen, der schließlich auch auf gesamtgesellschaftlicher Ebene zum Problem werden kann (Sobieraj, 2018). Ähnliche Konsequenzen lassen sich auch für andere Personengruppen erwarten, die etwa aufgrund ihrer sexuellen Orientierung, Religionszugehörigkeit oder Ethnie zu Opfern von Inzivilität und Hass werden.

Um herauszufinden, wie es sich für Betroffene anfühlt, das Ziel von Hassrede zu sein, konfrontierte Leets (2002) in einer Studie jüdische und homosexuelle Studierende mit Hasskommentaren, die ihre jeweilige soziale Identität ansprachen (antisemitisch bzw. homophob). Direkt nach der Konfrontation mit den Hasskommentaren zeigten die Studierenden vor allem emotionale Reaktionen, fühlten sich alleine, gekränkt oder schockiert.

Doch auch langfristige Auswirkungen auf Einstellungen und das eigene Verhalten wurden von den Teilnehmer*innen für wahrscheinlich gehalten. Die Autorin schlussfolgert daher, „that the consequences of hate speech might be similar [...] to the effects experienced by recipients of other kinds of traumatic experiences“ (S. 354). Inzivile Kommunikation mag somit „lediglich“ mit Wirkungen auf der Mikroebene beginnen, ist gerade in der Summe aber auch auf der Makroebene ein gravierendes Problem. Dies wirft die Frage auf, wie sich Inzivilität verhindern oder zumindest eindämmen lässt.

- 1 Studien im Kontext von Online-Hassrede sprechen dafür, dass solche offenen, „in your face“ Formen von Inzivilität schlichtweg auch leichter erkannt werden können als eher implizite Varianten (Ben-David & Matamoros-Fernández, 2016; Borgeson & Valeri, 2004).
- 2 Vertrauensverluste durch Inzivilität sind mutmaßlich jedoch nicht nur für den Journalismus zu beklagen. In einer finnischen Studie mit jugendlichen Facebook-Nutzer*innen (Näsi, Räsänen, Hawdon, Holkeri, & Oksanen, 2015) konnte ein Zusammenhang zwischen der Wahrnehmung von Online-Inzivilität (hier: Hassrede) und dem Vertrauen in soziale Gruppen festgestellt werden: Nutzer*innen, die sich daran erinnerten, online mit Hass konfrontiert worden zu sein, gaben an, weniger Vertrauen in andere Menschen zu haben.

5.
Gegen-
maßnahmen:
Wie lässt sich
Inzivilität in
sozialer,
Online-Medien
verhindern?

5. Gegenmaßnahmen: Wie lässt sich Inzivilität in sozialen Online-Medien verhindern?

Inzivilität in sozialen Online-Medien kann sehr explizit sein – etwa, wenn sie Gewaltandrohungen enthält oder von der Verwendung verfassungsfeindlicher Symbole begleitet wird. Das NetzDG macht es an dieser Stelle möglich, dass *repressive Strategien* wie Löschungen einzelner Kommentare oder das Blockieren von konkreten Nutzer*innen eingesetzt werden können. Allerdings zeigt die empirische Realität, dass Inzivilität vielfältige Ausdrucksformen annehmen kann, die sich häufig unter der Schwelle strafrechtlicher Relevanz bewegen. Konkrete Gegenmaßnahmen sind daher vor allem in *präventiven Strategien* zu sehen und nicht ausschließlich in der Repression von Inzivilität.

Es lassen sich verschiedene präventive Maßnahmen voneinander abgrenzen, wobei wir in diesem Beitrag eine bewusste Auswahl treffen und vorrangig auf die Aspekte (1) *Community Management* und Moderation von inzivilen Kommentaren und weiteren Kommunikaten, (2) das Ausüben von Gegenrede, (3) den Einsatz von Gegenbotschaften sowie (4) die Förderung von Medienkompetenz eingehen werden.¹

Insbesondere im Kontext von Nachrichten-Webseiten und den SNS-Profilen von Medien- und Nachrichten Anbietern wurde die Rolle von (1) *Community Management* und Moderation intensiv diskutiert. *Community Management* meint zunächst ganz allgemein das Übersehen bzw. Überwachen von Diskussionskanälen und -plattformen, „but also actively asking for contributions, encouraging users to contribute, and managing and moderating online discussions“ (Bakker, 2014, S. 598). Moderation kann somit als spezifische Teilform eines übergeordneten *Community Managements* verstanden werden, bei der die Prozesse oder Inhalte von Online-Diskussionen organisiert und reguliert werden

(Ziegele & Jost, 2016, S. 4; siehe auch Frischlich, Boberg, & Quandt, 2019). Idealtypisch lassen sich dabei drei Formen von Moderation unterscheiden, die sich im Grad der Beteiligung von Medienanbietern und Nutzer*innen unterscheiden (siehe nachfolgend Ziegele & Jost, 2016, S. 4–6): *Kollaborative Moderation* setzt auf eine starke Beteiligung der Diskussionsteilnehmer*innen und ermöglicht diesen beispielsweise, inzivile Kommentare zu *flaggen*, d.h., als unangemessen zu kennzeichnen und den Anbietern zu melden (ausführlich wird *Flagging* bei Naab, Kalch, & Meitz, 2018 diskutiert). Ein Beitrag, der von vielen Nutzer*innen *Flags* erhalten hat, kann dann je nach Einstellungen der jeweiligen Diskussionsplattform automatisch entfernt oder zur manuellen Prüfung weitergeleitet werden. Daneben gibt es die *inhaltliche Moderation*, bei der professionelle Moderator*innen ohne explizite Angabe von Gründen Kommentare entfernen, die sie als unangemessen empfinden (bzw. von der Organisation als unangemessen definiert wurden, z.B. in sogenannten „Netiquetten“). Zu unterscheiden sind dabei Formen der *Pre-* und *Post-Moderation*. Bei der *Pre-Moderation* werden Kommentare gesichtet, bevor sie veröffentlicht werden, während bei der *Post-Moderation* erst nach der Veröffentlichung festgelegte Prüfmechanismen zum Einsatz kommen (Ksiazek, 2015, S. 560). Zuletzt finden sich in der Praxis auch Formen der *interaktiven Moderation*, die sich dadurch auszeichnen, dass die Moderator*innen aktiv und sichtbar an den Diskussionen teilnehmen und mit den Nutzer*innen interagieren, wobei hier grob zwischen einer Helfer- (z.B. Bereitstellung von zusätzlichen Informationen; Komplimente für hilfreiche Beiträge) und Regulator-Rolle (z.B. Vermittlung bei Konflikten; Rückführung der Diskussionen zum eigentlichen Thema) differenziert werden kann.

Die zentrale Frage lautet an dieser Stelle, ob Moderation tatsächlich den gewünschten Erfolg bringt und in der Lage ist, Inzivilität in Online-Diskussionen zu verhindern. Eine Studie von Meyer & Carey (2014) zeigt diesbezüglich, dass eine aktive Form der Moderation dazu beiträgt, dass Nutzer*innen ein größeres (virtuelles) Gemeinschaftsgefühl verspüren, was sich wiederum positiv auf die allgemeine Diskussionskultur auswirken kann. Die Autoren raten vor diesem Hintergrund auch Journalist*innen, sich aktiv in Diskussionen einzuschalten und eine „noticeable presence in comment forums“ (ebd., S. 223) anzustreben. Dass sich eine solche Präsenz von Journalist*innen bei Diskussionen zu den von ihnen verfassten Nachrichtenbeiträgen positiv auswirken kann, zeigt sich auch in einer Studie von Stroud und Kolleg*innen (2015): Aktives *Involvement* von Reporter*innen war mit sichtbar weniger Inzivilität in den Kommentaren verbunden. Ähnlich findet auch Ksiazek (2018), dass Diskussionsbeiträge von Journalist*innen nicht nur mit einem Anstieg der Zahl an Nutzerkommentaren in Zusammenhang stehen, sondern auch deren Qualität zu verbessern scheinen. Des Weiteren offenbart seine Untersuchung, dass sowohl *Pre-* als auch *Post-Moderation* mit hochwertigeren (d.h. zivileren und/oder weniger feindseligen) Kommentaren assoziiert sind. Ziegele und Jost (2016) fokussieren indes auf Unterarten der interaktiven Moderation und finden mittels einer Experimentalstudie heraus, dass Beobachter*innen die Gesprächsatmosphäre einer Diskussion, in der ein Nachrichtenanbieter sachlich und höflich auf inzivile Kommentare reagiert, als positiver wahrnehmen als unmoderierte Diskussionen. Sarkastische oder satirische Kommentare leisten indes keinen positiven Beitrag, insbesondere mit Blick auf die Wahrnehmung des so kommunizierenden Anbieters. Grundsätzlich aber scheint Moderation – insbesondere jene, die als konstruktiv und sachlich charakterisiert werden kann – ein sinnvoller Ansatz zur Verminderung von Inzivilität zu sein.

Eine weitere Strategie zur Verhinderung (bzw. Verminderung) von Inzivilität ist **(2) das Ausüben von Gegenrede** (engl. *counter speech*). Die grundsätzliche Herangehensweise dieser primär im Kontext von Online-Hassrede eingesetzten

Gegenmaßnahme lässt sich mit dem Satz „the best remedy against hate speech is more speech“ (Delgado & Stefancic, 2004, S. 35) treffend zusammenfassen. Gegenrede zielt darauf, Hassrede (und deren negative Auswirkungen) zu überwinden, indem man sie einem Argumentationsprozess unterzieht (vgl. Sponholz, 2018). Ein Beispiel für eine deutsche Gegenrede-Kampagne ist die 2016 gegründete Facebook-Gruppe #ichbinhier (Kreißel, Ebner, Urban, & Guhl, 2018; Ley, 2018). Ziel der gegenwärtig knapp 45.000 Gruppenmitglieder umfassenden Initiative ist es, „zu einer anständige[n] Debattenkultur und zu einem vielfältigen Meinungsbild in den Kommentarspalten bei[zu]tragen“ (ichbinhier.eu/). Dazu sollen sich die Mitglieder aktiv an Facebook-Diskussionen beteiligen, die in den Kommentarspalten von reichweitenstarken Nachrichten Anbietern stattfinden und angesichts ihrer emotional besetzten oder kontroversen Thematik (z.B. Zuwanderung) inzivile Kommentare „anziehen“. Hierfür werden innerhalb der Facebook-Gruppe Aktionen ausgerufen, die die Mitglieder auffordern, der beobachteten Inzivilität mit respektvoller und sachlicher Gegenrede zu kontern. Allerdings scheinen sich nur wenige Mitglieder regelmäßig an den Aktionen zu beteiligen (Ley, 2018). Ein Grund dafür könnte der sogenannte Zuschauer Effekt (engl. *bystander effect*) sein. Dieser besagt, dass Augenzeug*innen eines Unfalls oder Verbrechens weniger wahrscheinlich eingreifen und helfen, je mehr andere Zuschauer*innen in die Situation involviert sind (Darley & Batson, 1973). Das heißt: Wenn schon jemand anders Gegenrede betrieben hat, sinkt die Intention, selbst Gegenrede auszuüben. Tatsächlich zeigt sich in einer aktuellen Experimentalstudie, dass die Bereitschaft zu Gegenrede bei einer hohen Zahl an *Bystandern* geringer ist und vorausgehende Reaktionen anderer die wahrgenommene Interventionsverantwortung der Nutzer*innen verringern (Leonhard, Rueß, Obermaier, & Reinemann, 2018). Allerdings kann der Schweregrad der Hassrede die Nutzer*innen dazu bewegen, ultimativ doch aktiv zu werden: Stärkere Hassrede führte eher dazu, dass die Befragten die Intention hatten, Gegenrede zu betreiben (ebd.). Insgesamt ist das Wissen darüber, wie effektiv Gegenrede zur (nachhaltigen) Bekämpfung von Inzivilität ist, aktuell jedoch

als gering einzustufen. Künftige Studien sollten die Frage nach der Wirksamkeit daher stärker in den Fokus der Aufmerksamkeit rücken.

Während Gegenrede von den Nutzer*innen sozialer Online-Medien initiiert und ausgelöst wird, werden unter **(3) Gegenbotschaften** professionell produzierte Medieninhalte verstanden, die in der wissenschaftlichen Auseinandersetzung daher häufig auch als *Counter-Narrative* bezeichnet werden (Briggs & Feve, 2013; Frischlich & Rieger, 2017; Yavuz, 2017). Unter Gegenbotschaften werden allgemein Videos, Texte oder auch *Memes* subsumiert, die sich gegen Diskriminierung und für Toleranz aussprechen. Dazu zählen etwa Geschichten, in denen ehemalige Hassredner*innen berichten, warum sie damit aufgehört haben sowie Narrative, die falsche Informationen oder ideologisch-verzerrte Weltbilder entkräften oder widerlegen (Rieger, Morten, & Frischlich, 2017). Ihre Verbreitung hat als Maßnahme in der Zivilgesellschaft viel Aufmerksamkeit bekommen. Eine der Hauptideen hinter der Verbreitung von Gegenbotschaften ist es, Online-Inzivilität (oder gar Extremismus) in der gleichen Umgebung zu begegnen, in der sie auch verbreitet wird. Die Produzent*innen von Gegenbotschaften setzen folglich auf Viralität und Popularität, um eine große Reichweite zu erzielen. Die Forschung zur Wirkung von derartigen Gegenbotschaften zeigt bislang uneinheitliche Befunde: So animieren etwa Gegenbotschaften, die in Form von Videos auf YouTube verbreitet werden (bspw. die Kampagne #whatIs von der Bundeszentrale für politische Bildung), nicht nur zu positiven Reaktionen oder Toleranz, sondern provozieren auch Hassrede in den zugehörigen Kommentarsektionen (Ernst u. a., 2017). Gegenbotschaften scheinen in sozialen Online-Medien also gewissermaßen ein Auslöser für „neue“ Inzivilität zu sein. Andererseits gibt es auch Hinweise, dass Gegenbotschaften die gewünschte positive Wirkung erzielen, wobei insbesondere die Narrativität der Botschaften als Grund für deren Effektivität ausgemacht werden konnte (Frischlich, Rieger, Morten, & Bente, 2018). Je mehr Gegenbotschaften den gängigen Prinzipien „guter“ Filme oder Kampagnen folgen und eine kohärente Geschichte erzählen, desto eher können sie auch eine positive Wirkung erzielen.

Dadurch erhöht sich zudem die Wahrscheinlichkeit, dass ein Video mit einer Gegenbotschaft weitergeleitet wird (Frischlich, Morten, Rieger, & Bente, 2017). Gegenbotschaften wurden vorrangig im Kontext der Extremismus-Prävention diskutiert, sind aber insbesondere aus einer übergeordneten Perspektive auch für die (Nicht-)Entstehung von Inzivilität relevant. Indem sie Toleranz für gesellschaftliche Gruppen schaffen, können sie insbesondere langfristig auch zu einer Verbesserung der Diskussions- und Debattenkultur beitragen.

Eine ähnliche übergeordnete Relevanz nimmt **(4) die Förderung von Medienkompetenz** ein, insbesondere für und bei jugendliche(n) Nutzer*innen sozialer Online-Medien. Denn selbst bei der Nutzung von eigentlich harmlosen Inhalten können die Empfehlungsalgorithmen sozialer Netzwerke dazu beitragen, dass (junge) Nutzer*innen ohne ihr Zutun mit problematischen und hasserfüllten Inhalten und Diskussionen konfrontiert werden (Schmitt, Rieger, Rutkowski, & Ernst, 2018). In solchen Situationen braucht es Wissen und Fähigkeiten, mit medial vermittelter Inzivilität umzugehen. Medienkompetenz lässt sich in die drei übergeordneten Bereiche der Sach-, Selbst- und Sozialkompetenz einteilen (siehe z.B. Pfaff-Rüdiger, Riesmeyer, & Kümpel, 2012; Riesmeyer, Pfaff-Rüdiger, & Kümpel, 2016), wobei im Themenfeld der inzivilen Kommunikation insbesondere die zwei letztgenannten Bereiche eine Rolle spielen dürften. So müssen Nutzer*innen nicht nur in der Lage sein, Inzivilität zu identifizieren und kritisch zu evaluieren (reflexive [Selbst-]Kompetenz) um gefühlsmäßig angemessen darauf zu reagieren (emotionale [Selbst-]Kompetenz), sondern insbesondere auch über interaktive und ethisch-moralische Sozialkompetenzen verfügen, d.h. produktiv partizipieren und problematische (Diskussions-)Situationen erkennen können. Allerdings ist auch eine gewisse Sachkompetenz vonnöten: Nur wer weiß, dass und wie man inzivile Kommentare melden oder *flaggen* kann, Beratungsstellen oder Initiativen wie kennt, ist in der Lage, gegen Inzivilität vorzugehen. Pädagogische Initiativen und Lehrkräfte sollten folglich sowohl das nötige Handwerkszeug für Diskussionen in der Online-Welt bereitstellen, als auch eine generelle Auseinandersetzung mit der heutigen

Sprach- und Debattenkultur vorantreiben. Im Bereich der Extremismusprävention wurde bisher vor allem versucht, die Medienkritikfähigkeit, also eine Form der reflexiven Selbstkompetenz, zu fördern (siehe z.B. Rieger, Ernst, u. a., 2017; Schmitt u. a., 2019). In einer ersten Evaluation entsprechend entwickelter Unterrichtseinheiten konnte gezeigt werden, dass vor allem das Bewusstsein für und das Wissen über extremistische Online-Inhalte (d.h., eine generelle *Awareness*) gestärkt werden konnten (Schmitt u. a., 2018).

1 Nicht im Detail diskutieren werden wir das Prinzip der direkten Ansprache/Intervention. Direkte Ansprache beschreibt ein besonders ressourcenintensives Vorgehen, bei dem proaktiv auf jene Nutzer*innen sozialer Online-Medien zugegangen wird, die sich (wiederholt) inzivil, hasserfüllt oder extremistisch geäußert haben. Ihnen werden – häufig durch ehemalige Täter*innen, Opfer oder professionelle Berater*innen – Gesprächsoptionen aufgezeigt sowie Unterstützung angeboten. Ein Beispiel für einen solchen Ansatz der sozialen Online-Arbeit ist das „One to One“-Programm des Institute for Strategic Dialogue (ISD), bei dem Nutzer*innen mit extremistischen Ansichten in einen konstruktiven Dialog verwickelt werden (Frenett & Dow, 2015; Institute for Strategic Dialogue, 2019).

5. Gegenmaßnahmen: Wie lässt sich Inzivilität in sozialen Online-Medien verhindern?

6. Fazit und Ausblick

6. Fazit und Ausblick

Soziale Online-Medien sind aus dem Alltag nicht mehr wegzudenken und in der Lage, eine Vielzahl an sozial-integrativen, persönlichen und informationsbezogenen Bedürfnissen zu erfüllen. Sie bieten nicht zuletzt einen Raum, in dem auch „normale“ Bürger*innen Themen von öffentlicher Relevanz diskutieren und laufende Debatten beobachten können. Neben den vielfältigen positiven Potenzialen, die sich aus den gesteigerten Möglichkeiten für Partizipation und Teilhabe ergeben, sind in sozialen Online-Medien jedoch auch die Schattenseiten dieser Entwicklung ersichtlich. Viele Online-Diskussionen sind von Inzivilität durchzogen und die Konfrontation mit Hass, Beleidigungen und Pöbeleien scheint mehr Regel denn Ausnahme zu sein.

Angesichts dieser Entwicklung hat der vorliegende Beitrag die negativen Aspekte eines Wandels der Sprach- und Debattenkultur in sozialen Online-Medien in den Blick genommen. Konkret haben wir mittels eines narrativen Literaturüberblicks sowohl mögliche Ursachen (der Wahrnehmung) einer inzivilen Sprach- und Debattenkultur als auch Wirkungen auf die Nutzer*innen sozialer Online-Medien diskutiert. Darauf aufbauend wurden abschließend mögliche Gegenmaßnahmen vorgestellt, die zu einer Verminderung von Inzivilität – und somit einer Verbesserung der diskursiven Qualität – beitragen können.

Die **Ursachen** für die Wahrnehmung von mehr Inzivilität in sozialen Online-Medien wurden auf Basis bisheriger Befunde in vier übergeordnete Bereiche unterteilt. Zunächst ist davon auszugehen, dass (1) die gesteigerte Sichtbarkeit und öffentliche Zugänglichkeit von Debatten und Diskursen – insbesondere in Kombination mit (2) vereinfachten Möglichkeiten zur schnellen und weiten Verbreitung von Online-Inhalten – dazu geführt hat, dass Nutzer*innen schlichtweg mehr Gelegenheiten haben, gesellschaftliche Kommunikation (und somit auch inzivile Formen dieser) wahrzunehmen. Insbesondere soziale

Netzwerkseiten wie Facebook und Twitter haben außerdem die Weiterverbreitung von Inhalten und Diskussionen drastisch erleichtert und tragen durch ihre Empfehlungsalgorithmen dazu bei, dass sich vor allem emotional aufgeladene und polarisierende Beiträge weit und schnell verbreiten. Studien zu den (3) veränderten Kommunikationsbedingungen und (Gruppen-)Normen in sozialen Online-Medien deuten des Weiteren darauf hin, dass Anonymität, die Unsichtbarkeit der Diskussionsteilnehmer*innen sowie das Fehlen von regulativen sozialen Hinweisreizen Hemmungen abbauen und so zu einem Mehr an inziviler Kommunikation führen können. Wird Inzivilität bei anderen beobachtet, steigt zudem die Wahrscheinlichkeit, sich selbst entsprechender Ausdrucksformen zu bedienen. Zuletzt sind auch (4) motivationale Aspekte und Persönlichkeitsmerkmale als Ursachen von Inzivilität zu berücksichtigen. Eine Kenntnis über individuelle Prädispositionen und Motive ist daher ebenso unerlässlich wie der Blick auf die zuvor diskutierten Ursachenkataloge.

Die **Wirkungen** von Inzivilität hängen in einem ersten Schritt davon ab, ob beobachtete Online-Diskussionen und Debatten überhaupt als inzivil wahrgenommen werden. Insbesondere bei der Androhung von Gewalt und Beleidigungen scheint diesbezüglich bei den meisten Beobachter*innen Einigkeit zu bestehen. Allerdings zeigt sich auch, dass Frauen und Personen, die sich als mitfühlend und kooperativ beschreiben, tendenziell mehr Inzivilität wahrnehmen. Mit Blick auf die beobachtenden Nutzer*innen zeigen sich Einflüsse von inziviler Kommunikation insbesondere auf (1) medien- bzw. journalismusbezogene Perzeptionen, (2) Emotionen und (aggressive) Verhaltensintentionen sowie (3) das eigene Kommunikationsverhalten. Die Wirkungen sind bisherigen Befunden zufolge tatsächlich in aller Regel *negativ*. Wer inzivile Kommentare liest, beurteilt etwa die kommentierten Inhalte (z.B. Nachrichtenbeiträge) schlechter und hat verzerrte Vorstellungen über

die Relevanz der behandelten Themen oder die Glaubwürdigkeit journalistischer Angebote. Inzivilität kann zudem negative Emotionen evozieren und aggressive Gedanken auslösen sowie das eigene Kommunikationsverhalten negativ beeinflussen. Besonders schwerwiegend sind Wirkungen jedoch für die von Inzivilität Betroffenen. Bei Diskussionen von Nachrichtenthemen sind vor allem Journalist*innen die Opfer inziviler Kommunikation, was sich nicht nur auf die hinter den Artikeln stehenden Individuen, sondern auch auf deren redaktionelle Arbeit auswirken kann. Daneben werden Frauen, religiöse und ethnische Minderheiten sowie Menschen mit einer individuellen sozialen Identität in Online-Medien häufig mit Hass und Inzivilität konfrontiert. Dies hat nicht nur emotionale Reaktionen zur Folge, sondern kann langfristig auch dazu führen, dass sich die Betroffenen teilweise oder gar vollständig aus öffentlichen Diskursen zurückziehen.

Die **Gegenmaßnahmen** können in repressive und präventive Strategien unterteilt werden, wobei letztere bei der Bekämpfung von Inzivilität – insbesondere langfristig – die wichtigere Rolle einnehmen dürften. (1) *Community Management* und Moderation, (2) das Ausüben von Gegenrede, (3) die Produktion und Distribution von Gegenbotschaften sowie (4) die Förderung von Medienkompetenz sind diesbezüglich von besonderer Relevanz.

Die Zusammenstellung von Ursachen, Wirkungen und Gegenmaßnahmen im Kontext von Inzivilität in sozialen Online-Medien hat schließlich auch einige Lücken und Probleme in der wissenschaftlichen Auseinandersetzung offenbart. Diese beginnen bereits bei der Definition von Inzivilität: So gibt es enorm viele Subdimensionen und Erscheinungsformen von inziviler Kommunikation, die auch dieser Beitrag mutmaßlich nicht vollständig erfassen konnte. Schwierig ist ferner, dass selbst die Verwendung des gleichen Begriffes (z.B. *Flaming*, *Hate Speech*) nicht bedeutet, dass auch tatsächlich das identische Phänomen untersucht wird. Allerdings haben all diese Erscheinungsformen von Inzivilität auch einen gemeinsamen Kern und häufig ähnliche (negative) Auswirkungen auf Beobachter*innen und Betroffene. Aus unserer

Sicht lohnt bei der Untersuchung von jeglichen Formen von Inzivilität daher immer auch ein Blick über den Tellerrand spezifischer Begrifflichkeiten hinweg. Mit einer entsprechenden Offenheit ließe sich zudem eine terminologisch bedingte „Abschottung“ einzelner Forschungsbereiche vermeiden, die den Erkenntnisgewinn eher verhindert denn stimuliert.

Die schiere Menge an in sozialen Online-Medien anfallenden Inhalten, Kommunikaten und Debatten stellt die Forschung zu Inzivilität zudem vor methodische Herausforderungen. Wer Aussagen über generelle Tendenzen treffen will, kann hier unmöglich nur auf manuelle Formen der Analyse zurückgreifen. Zunehmend werden daher *Machine-Learning*-Techniken oder andere Ansätze zur automatischen Detektion von Inzivilität eingesetzt. Die Konfiguration entsprechender *Tools* ist jedoch alles andere als voraussetzungslos (siehe z.B. Davidson, Warmesley, Macy, & Weber, 2017; Saleem, Dillon, Benesch, & Ruths, 2017). So stellt sich etwa im Bereich Hassrede die Frage, wann ein Inhalt „nur“ beleidigend ist und wann man davon ausgehen kann, dass es sich um Hass gegen gesellschaftliche Gruppen handelt. Methodische Probleme ergeben sich darüber hinaus durch die Aneignung von als Beleidigung genutzten Ausdrücken durch die davon betroffenen Gruppen. Im US-amerikanischen Raum etwa nutzen viele Afroamerikaner*innen das Wort „Nigger“ nicht als Schimpfwort, sondern als freundschaftliche Anrede oder um Zugehörigkeit zur eigenen *Ingroup* zu signalisieren, was schlüsselwortbasierte Ansätze der Identifikation von Hassrede an ihre Grenzen bringt (Hasanuzzaman, Dias, & Way, 2017). Des Weiteren wird künftig auch die Analyse von bildbasierten Formen von Inzivilität immer wichtiger werden. So wird etwa Instagram derzeit als „the Internet’s new home for hate“ (Lorenz, 2019) diskutiert, da Inzivilität und Hassrede hier in besonderem Maße zu florieren scheinen. Diese problematische Entwicklung muss aus unserer Sicht auch durch entsprechende Wirkungsstudien flankiert werden, da davon auszugehen ist, dass Videos und (bildbasierte) *Memes* andere Effekte auf die Nutzer*innen haben als textbasierte Formen inziviler Kommunikation.

Ein Aspekt, der bislang weitgehend unerwähnt geblieben ist, ist die Beziehung zwischen Inzivilität in der virtuellen und „realen“ Welt. Am Beispiel von antimuslimischer Hassrede zeigen etwa Awan und Zempi (2016), dass die Grenze zwischen Online-Hassrede und „Offline“-Bedrohungen (oder gar Angriffen) fließend ist. Insbesondere für die Opfer fühlen sich die Konsequenzen gleichermaßen real an – unabhängig davon, wo Hass erfahren wird. Doch auch für die Beobachter*innen von Inzivilität kann man mit *Spillover*-Effekten rechnen. Für finnische Jugendliche finden Näsi und Kolleg*innen (2015), dass die Konfrontation mit Online-Hassrede auf Facebook das Vertrauen in andere Menschen ganz generell verringert: Wer mit inziviler Kommunikation im Internet konfrontiert ist, scheint das daraus resultierende negative Menschenbild also auch auf die „echte“ Welt zu übertragen. Inzivilität in sozialen Online-Medien ist somit fraglos ein Problem, das aus gesamtgesellschaftlicher Sicht adressiert werden muss.

Die Autorinnen

Prof. Dr. Diana Rieger lehrt am Institut für Kommunikationswissenschaft und Medienforschung der Ludwig-Maximilians-Universität München.

Dr. Anna Sophie Kümpel ist wissenschaftliche Mitarbeiterin am Institut für Kommunikationswissenschaft und Medienforschung der Ludwig-Maximilians-Universität München.

Literaturverzeichnis

- A** Adams, C. (2018). "They go for gender first". The nature and effect of sexist abuse of female technology journalists. *Journalism Practice*, 12(7), S. 850–869.
- 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), S. 373–387.
- 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), S. 156–168.
- Awan, I. (2014). Islamophobia and Twitter: A typology of online hate against Muslims on social media. *Policy & Internet*, 6(2), S. 133–150.
- Awan, I. (2016). Islamophobia on social media: A qualitative analysis of the Facebook's walls of hate. *International Journal of Cyber Criminology*, 10(1), S. 1–20.
- Awan, I., & Zempi, I. (2016). The affinity between online and offline anti-Muslim hate crime: Dynamics and impacts. *Aggression and Violent Behavior*, 27, S. 1–8.
- B** Bakker, P. (2014). Mr. Gates returns. Curation, community management and other new roles for journalists. *Journalism Studies*, 15(5), S. 596–606.
- Baym, N. K., & boyd, danah. (2012). Socially mediated publicness: An introduction. *Journal of Broadcasting & Electronic Media*, 56(3), S. 320–329.
- Ben-David, A., & Matamoros-Fernández, A. (2016). Hate Speech and covert discrimination on social media: Monitoring the Facebook pages of extreme-right political parties in Spain. *International Journal of Communication*, 10, S. 1167–1193.
- Berg, J. (2016). The impact of anonymity and issue controversiality on the quality of online discussion. *Journal of Information Technology & Politics*, 13(1), S. 37–51.
- Berger, J. (2011). Arousal increases social transmission of information. *Psychological Science*, 22(7), S. 891–893.
- Binns, A. (2017). Fair game? Journalists' experiences of online abuse. *Journal of Applied Journalism & Media Studies*, 6(2), S. 183–206.
- Blazak, R. (2009). Toward a working definition of hate groups. In B. Perry (Hrsg.), *Hate Crimes. Understanding and Defining Hate Crime* (Bd. 1, S. 133–162). Westport, CT: Praeger.

Blom, R., Carpenter, S., Bowe, B. J., & Lange, R. (2014). Frequent contributors within U.S. newspaper comment forums: An examination of their civility and information value. *American Behavioral Scientist*, 58(10), S. 1314–1328.

Borgeson, K., & Valeri, R. (2004). Faces of hate. *Journal of Applied Sociology*, 21(2), S. 99–111.

Briggs, R., & Feve, S. (2013). *Review of programs to counter narratives of violent extremism*. Abgerufen von <https://www.counterextremism.org/resources/details/id/444/review-of-programs-to-counter-narratives-of-violent-extremism-what-works-and-what-are-the-implications-for-government> am 17.04.2019.

Brown, A. (2018). What is so special about online (as compared to offline) hate speech? *Ethnicities*, 18(3), S. 297–326.

Buckels, E. E., Trapnell, P. D., & Paulhus, D. L. (2014). Trolls just want to have fun. *Personality and Individual Differences*, 67, S. 97–102.

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

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

Chen, G. M., Pain, P., Chen, V. Y., Mekelburg, M., Springer, N., & Troger, F. (2018). 'You really have to have a thick skin': A cross-cultural perspective on how online harassment influences female journalists. *Journalism*. Advance Online Publication.

Cheng, J., Bernstein, M., Danescu-Niculescu-Mizil, C., & Leskovec, J. (2017). Anyone can become a troll: Causes of trolling behavior in online discussions. *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, S. 1217–1230.

Chetty, N., & Alathur, S. (2018). Hate speech review in the context of online social networks. *Aggression and Violent Behavior*, 40, S. 108–118.

Christopherson, K. M. (2007). The positive and negative implications of anonymity in Internet social interactions: "On the Internet, Nobody Knows You're a Dog". *Computers in Human Behavior*, 23(6), S. 3038–3056.

Cicchirillo, V., Hmielowski, J., & Hutchens, M. (2015). The mainstreaming of verbally aggressive online political behaviors. *Cyberpsychology, Behavior, and Social Networking*, 18(5), S. 253–259.

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), S. 658–679.

- Costello, M., Hawdon, J., & Ratliff, T. N. (2017). Confronting online extremism: The effect of self-help, collective efficacy, and guardianship on being a target for hate speech. *Social Science Computer Review*, 35(5), S. 587–605.
- Craker, N., & March, E. (2016). The dark side of Facebook®: The Dark Tetrad, negative social potency, and trolling behaviours. *Personality and Individual Differences*, 102, S. 79–84.
- D** Darley, J. M., & Batson, C. D. (1973). „From Jerusalem to Jericho“: A study of situational and dispositional variables in helping behavior. *Journal of Personality and Social Psychology*, 27(1), S. 100–108.
- Davidson, T., Warmesley, D., Macy, M., & Weber, I. (2017). Automated hate speech detection and the problem of offensive language. *Proceedings of the 11th International Conference on Web and Social Media*.
- Delgado, R., & Stefancic, J. (2004). *Understanding words that wound*. Boulder, CO: Westview Press.
- DeVito, M. A. (2017). From editors to algorithms. A values-based approach to understanding story selection in the Facebook news feed. *Digital Journalism*, 5(6), S. 753–773.
- Dohle, M. (2018). Recipients' assessment of journalistic quality. Do online user comments or the actual journalistic quality matter? *Digital Journalism*, 6(5), S. 563–582.
- Dutton, W. H. (1996). Network rules of order: Regulating speech in public electronic fora. *Media, Culture & Society*, 18(2), S. 269–290.
- E** Eilders, C. (1997). *Nachrichtenfaktoren und Rezeption. Eine empirische Analyse zur Auswahl und Verarbeitung politischer Information*. Opladen: Westdeutscher Verlag.
- Erjavec, K., & Kovačič, M. P. (2012). “You don’t understand, this is a new war!” Analysis of hate speech in news web sites’ comments. *Mass Communication and Society*, 15(6), S. 899–920.
- Ernst, J., Schmitt, J. B., Rieger, D., Beier, A. K., Vorderer, P., Bente, G., & Roth, H.-J. (2017). Hate beneath the counter speech? A qualitative content analysis of user comments on YouTube related to counter speech videos. *Journal for Deradicalization*, (10), S. 1–49.
- European Federation of Journalists. (2017, Oktober 31). There can be no press freedom where journalists work in fear, end impunity! Abgerufen von <https://european-journalists.org/blog/2017/10/31/there-can-be-no-press-freedom-where-journalists-work-in-fear-end-impunity/> am 17.04.2019.
- F** Fawzi, N. (2015). *Cyber-Mobbing: Ursachen und Auswirkungen von Mobbing im Internet* (2., durchgesehene Auflage). Baden-Baden: Nomos.

Fox, J., & Tang, W. Y. (2014). Sexism in online video games: The role of conformity to masculine norms and social dominance orientation. *Computers in Human Behavior*, 33, S. 314–320.

Fredheim, R., Moore, A., & Naughton, J. (2015). Anonymity and online commenting: The broken windows effect and the end of drive-by commenting. *Proceedings of the ACM Web Science Conference on WWW*.

Frees, B., & Koch, W. (2018). ARD/ZDF-Onlinestudie 2018: Zuwachs bei medialer Internetnutzung. *Media Perspektiven*, (9), S. 398–413.

Frenett, R., & Dow, M. (2015). *One to one online interventions - A pilot CVE methodology* (S. 1–27). Abgerufen von http://www.isdglobal.org/wp-content/uploads/2016/04/One2One_Web_v9.pdf am 17.04.2019.

Frischlich, L., Boberg, S., & Quandt, T. (2019). Comment sections as targets of dark participation? Journalists' evaluation and moderation of deviant user comments. *Journalism Studies*. Advance Online Publication.

Frischlich, L., Morten, A., Rieger, D., & Bente, G. (2017). Wirkung. In L. Frischlich, A. Morten, D. Rieger, G. Bente, & in Kooperation mit der Forschungsstelle Terrorismus/Extremismus des Bundeskriminalamtes (Hrsg.), *Videos gegen Extremismus? Counter-Narrative auf dem Prüfstand* (S. 81–140). Wiesbaden: Bundeskriminalamt.

Frischlich, L., & Rieger, D. (2017). Hass im Netz – Hass im Herzen? Die Wirkung rechtsextremistischer und islamistisch-extremistischer Online Propagandavideos und mögliche Gegenangebote im Netz. *The Inquisitive Mind*, 2. Abgerufen von <http://de.in-mind.org/article/hass-im-netz-hass-im-herzen-die-wirkung-rechtsextremistischer-und-islamistisch> am 17.04.2019.

Frischlich, L., Rieger, D., Morten, A., & Bente, G. (2018). The power of a good story: Narrative persuasion in extremist propaganda and videos against violent extremism. *International Journal of Conflict and Violence (IJCV)*, 12, S. 1–16.

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

Glaser, J., Dixit, J., & Green, D. P. (2002). Studying hate crime with the internet: What makes racists advocate racial violence? *Journal of Social Issues*, 58(1), S. 177–193.

Glaser, M., Herzog, H., Özkilic, M., & Schindler, F. (2016). *Jugendschutz im Internet - Ergebnisse der Recherchen und Kontrollen. Bericht 2015*. Mainz: Jugendschutz.net.

Glaser, M., Herzog, H., Özkilic, M., & Schindler, F. (2018). *Jugendschutz im Internet - Risiken und Handlungsbedarf. Bericht 2017*. Mainz: Jugendschutz.net.

Gollatz, K., Riedl, M. J., & Pohlmann, J. (2018, August 9). *Removals of online hate speech in numbers*. Abgerufen von <https://doi.org/10.5281/zenodo.1342325> am 17.04.2019.

- H Han, S.-H., & Brazeal, L. M. (2015). Playing nice: Modeling civility in online political discussions. *Communication Research Reports*, 32(1), S. 20–28.
- 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), S. 1–12.
- Hasanuzzaman, M., Dias, G., & Way, A. (2017). Demographic word embeddings for racism detection on Twitter. *Proceedings of the 8th International Joint Conference on Natural Language Processing*, S. 926–936. Taipei, Taiwan.
- Hasell, A., & Weeks, B. E. (2016). Partisan provocation: The role of partisan news use and emotional responses in political information sharing in social media. *Human Communication Research*, 42(4), S. 641–661.
- Herbst, S. (2010). *Rude democracy: Civility and incivility in American politics*. Philadelphia, PA, USA: Temple University Press.
- Hmielowski, J. D., Hutchens, M. J., & Cicchirillo, V. J. (2014). Living in an age of online incivility: Examining the conditional indirect effects of online discussion on political flaming. *Information, Communication & Society*, 17(10), S. 1196–1211.
- Ho, S. S., & McLeod, D. M. (2008). Social-psychological influences on opinion expression in face-to-face and computer-mediated communication. *Communication Research*, 35(2), S. 190–207.
- Hölig, S., & Hasebrink, U. (2018). *Reuters Institute Digital News Report 2018 – Ergebnisse für Deutschland* (Nr. 44). Hamburg, Deutschland: Hans-Bredow-Institut.
- Houston, J. B., Hansen, G. J., & Nisbett, G. S. (2011). Influence of user comments on perceptions of media bias and third-person effect in online news. *Electronic News*, 5(2), S. 79–92.
- Hsueh, M., Yogeewaran, K., & Malinen, S. (2015). "Leave your comment below": Can biased online comments influence our own prejudicial attitudes and behaviors? *Human Communication Research*, 41(4), S. 557–576.
- Hutchens, M. J., Cicchirillo, V. J., & Hmielowski, J. D. (2015). How could you think that?!?: Understanding intentions to engage in political flaming. *New Media & Society*, 17(8), S. 1201–1219.
- I Institute for Strategic Dialogue. (2019). One to One. Abgerufen von <https://www.isdglobal.org/programmes/communications-technology/one-to-one/> am 17.04.2019.
- J Jane, E. A. (2015). Flaming? What flaming? The pitfalls and potentials of researching online hostility. *Ethics and Information Technology*, 17(1), S. 65–87.
- Johnen, M., Jungblut, M., & Ziegele, M. (2018). The digital outcry: What incites participation behavior in an online firestorm? *New Media & Society*, 20(9), S. 3140–3160.

Joinson, A. N. (2007). Disinhibition and the Internet. In J. Gackenbach (Hrsg.), *Psychology and the Internet* (S. 75–92).

- K** Kaakinen, M., Oksanen, A., & Räsänen, P. (2018). Did the risk of exposure to online hate increase after the November 2015 Paris attacks? A group relations approach. *Computers in Human Behavior*, *78*, S. 90–97.

Kenski, K., Coe, K., & Rains, S. A. (2017). Perceptions of uncivil discourse online: An examination of types and predictors. *Communication Research*. Advance Online Publication.

KhosraviNik, M., & Esposito, E. (2018). Online hate, digital discourse and critique: Exploring digitally-mediated discursive practices of gender-based hostility. *Lodz Papers in Pragmatics*, *14*(1), S. 45–68.

Kiesler, S., Siegel, J., & McGuire, T. W. (1984). Social psychological aspects of computer-mediated communication. *American Psychologist*, *39*(10), S. 1123–1134.

Klein, A. (2012). Slipping racism into the mainstream: A theory of information laundering. *Communication Theory*, *22*(4), S. 427–448.

Kreißel, P., Ebner, J., Urban, A., & Guhl, J. (2018). *Hass auf Knopfdruck. Rechtsextreme Trollfabriken und das Ökosystem koordinierter Hasskampagnen im Netz*. London: Institute for Strategic Dialogue.

Ksiazek, T. B. (2015). Civil interactivity: How news organizations' commenting policies explain civility and hostility in user comments. *Journal of Broadcasting & Electronic Media*, *59*(4), S. 556–573.

Ksiazek, T. B. (2018). Commenting on the news. Explaining the degree and quality of user comments on news websites. *Journalism Studies*, *19*(5), S. 650–673.

Ksiazek, T. B., & Springer, N. (2018). User comments in digital journalism: Current research and future directions. In S. A. Eldridge & B. Franklin (Hrsg.), *The Routledge Handbook of Developments in Digital Journalism Studies*. London: Routledge.

Kümpel, A. S. (2019). *Nachrichtenrezeption auf Facebook. Vom beiläufigen Kontakt zur Auseinandersetzung*. Wiesbaden: Springer VS.

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).

Kümpel, A. S., & Springer, N. (2016). Commenting quality. Effects of user comments on perceptions of journalistic quality. *Studies in Communication and Media (SCM)*, *5*(3), S. 353–366.

Kümpel, A. S., & Unkel, J. (im Druck). Negativity wins at last: How presentation order and valence of user comments affect perceptions of journalistic quality. *Journal of Media Psychology*.

- L** Lapidot-Lefler, N., & Barak, A. (2012). Effects of anonymity, invisibility, and lack of eye-contact on toxic online disinhibition. *Computers in Human Behavior*, 28(2), S. 434–443.
- Leets, L. (2002). Experiencing hate speech: Perceptions and responses to anti-semitism and antigay speech. *Journal of Social Issues*, 58(2), S. 341–361.
- Leets, L., & Giles, H. (1999). Harmful speech in intergroup encounters: An organizational framework for communication research. *Annals of the International Communication Association*, 22(1), S. 91–137.
- Leonhard, L., Rueß, C., Obermaier, M., & Reinemann, C. (2018). Perceiving threat and feeling responsible. How severity of hate speech, number of bystanders, and prior reactions of others affect bystanders' intention to counterargue against hate speech on Facebook. *Studies in Communication and Media (SCM)*, 7(4), S. 555–579.
- Ley, H. (2018). *#ichbinhier: Zusammen gegen Fake News und Hass*. Köln: DuMont.
- Lorenz, T. (2019, März 21). Instagram is the Internet's new home for hate. *The Atlantic*. Abgerufen von <https://www.theatlantic.com/technology/archive/2019/03/instagram-is-the-internets-new-home-for-hate/585382/> am 17.04.2019.
- M** Matzner, T. (2016). Anonymität. In J. Heesen (Hrsg.), *Handbuch Medien- und Informationsethik* (S. 248–254). Stuttgart: J.B. Metzler.
- McNamee, L. G., Peterson, B. L., & Peña, J. (2010). A call to educate, participate, invoke and indict: Understanding the communication of online hate groups. *Communication Monographs*, 77(2), S. 257–280.
- Meibauer, J. (2013). Hassrede - von der Sprache zur Politik. In J. Meibauer (Hrsg.), *Hassrede/Hate Speech. Interdisziplinäre Beiträge zu einer aktuellen Diskussion* (S. 1–16). Gießen: Gießener Elektronische Bibliothek.
- Meyer, H. K., & Carey, M. C. (2014). In moderation. Examining how journalists' attitudes toward online comments affect the creation of community. *Journalism Practice*, 8(2), S. 213–228.
- Moor, P. J., Heuvelman, A., & Verleur, R. (2010). Flaming on YouTube. *Computers in Human Behavior*, 26(6), S. 1536–1546.
- Muddiman, A. (2017). Personal and public levels of political incivility. *International Journal of Communication*, 11, S. 3182–3202.
- N** Naab, T. K., Kalch, A., & Meitz, T. G. (2018). Flagging uncivil user comments: Effects of intervention information, type of victim, and response comments on bystander behavior. *New Media & Society*, 20(2), S. 777–795.

- Näsi, M., Räsänen, P., Hawdon, J., Holkeri, E., & Oksanen, A. (2015). Exposure to online hate material and social trust among Finnish youth. *Information Technology & People*, 28(3), S. 607–622.
- Neubaum, G., & Krämer, N. C. (2017). Opinion climates in social media: Blending mass and interpersonal communication. *Human Communication Research*, 43(4), S. 464–476.
- Neuberger, C. (2017). Soziale Medien und Journalismus. In J.-H. Schmidt & M. Tad-dicken (Hrsg.), *Handbuch Soziale Medien* (S. 101–127). Wiesbaden: Springer VS.
- Nilsson, M. L., & Örnebring, H. (2016). Journalism under threat. Intimidation and harassment of Swedish journalists. *Journalism Practice*, 10(7), S. 880–890.
- O Obermaier, M., Hofbauer, M., & Reinemann, C. (2018). Journalists as targets of hate speech. How German journalists perceive the consequences for themselves and how they cope with it. *Studies in Communication and Media (SCM)*, 7(4), S. 499–524.
- Oksanen, A., Kaakinen, M., Minkkinen, J., Räsänen, P., Enjolras, B., & Steen-Johnsen, K. (2018). Perceived societal fear and cyberhate after the November 2015 Paris terrorist attacks. *Terrorism and Political Violence*. Advance Online Publication.
- Osatuyi, B. (2013). Information sharing on social media sites. *Computers in Human Behavior*, 29(6), S. 2622–2631.
- O'Sullivan, P. B., & Flanagin, A. J. (2003). Reconceptualizing 'flaming' and other problematic messages. *New Media & Society*, 5(1), S. 69–94.
- Owen, L. H. (2018, Januar 11). Facebook drastically changes News Feed to make it "good for people" (and bad for most publishers). Abgerufen von <http://www.niemanlab.org/2018/01/facebook-dramatically-changes-news-feed-to-make-it-good-for-people-and-bad-for-most-publishers/> am 17.04.2019.
- Owen, L. H. (2019, März 15). One year in, Facebook's big algorithm change has spurred an angry, Fox News-dominated — and very engaged! — News Feed. Abgerufen von <http://www.niemanlab.org/2019/03/one-year-in-facebooks-big-algorithm-change-has-spurred-an-angry-fox-news-dominated-and-very-engaged-news-feed/> am 17.04.2019.
- P Papacharissi, Z. (2004). Democracy online: Civility, politeness, and the democratic potential of online political discussion groups. *New Media & Society*, 6(2), S. 259–283.
- Pfaff-Rüdiger, S., Riesmeyer, C., & Kümpel, A. S. (2012). Media literacy and developmental tasks: A case study in Germany. *Media Studies*, 3(6), S. 42–56.
- Pfeffer, J., Zorbach, T., & Carley, K. M. (2014). Understanding online firestorms: Negative word-of-mouth dynamics in social media networks. *Journal of Marketing Communications*, 20(1–2), S. 117–128.

- Post, S., & Kepplinger, H. M. (2019). Coping with audience hostility. How journalists' experiences of audience hostility influence their editorial decisions. *Journalism Studies*. Advance Online Publication.
- Postmes, T., Spears, R., & Lea, M. (1998). Breaching or building social boundaries? SIDE-effects of computer-mediated communication. *Communication Research*, 25(6), S. 689–715.
- Preuß, M., Tetzlaff, F., & Zick, A. (2017). „Publizieren wird zur Mutprobe“. Studie zur Wahrnehmung von und Erfahrungen mit Angriffen unter Journalist_innen. Abgerufen von <https://mediendienst-integration.de/fileadmin/Dateien/Studie-hatespeech.pdf> am 17.04.2019.
- Prochazka, F., Weber, P., & Schweiger, W. (2018). Effects of civility and reasoning in user comments on perceived journalistic quality. *Journalism Studies*, 19(1), S. 62–78.
- Puryear, C., & Vandello, J. A. (2018). Inflammatory comments elicit less outrage when made in anonymous online contexts. *Social Psychological and Personality Science*. Advance Online Publication.
- R** Reicher, S. D., Spears, R., & Postmes, T. (1995). A social identity model of deindividuation phenomena. *European Review of Social Psychology*, 6(1), S. 161–198.
- Reinemann, C., Nienierza, A., Fawzi, N., Riesmeyer, C., & Neumann, K. (2019). *Jugend – Medien – Extremismus. Wo Jugendliche mit Extremismus in Kontakt kommen und wie sie ihn erkennen*. Wiesbaden: Springer VS.
- Rieger, D., Dippolt, J., & Appel, M. (im Druck). Trolle gibt es nicht nur im Märchen – Das Phänomen Trolling im Internet. In M. Appel (Hrsg.), *Die Psychologie des Postfaktischen: Über „Fake News“, „Lügenpresse“, „Clickbaits“ und Co.* Wiesbaden: Springer VS.
- Rieger, D., Ernst, J., Schmitt, J. B., Vorderer, P., Bente, G., & Roth, H.-J. (2017). Propaganda und Alternativen im Internet. Medienpädagogische Implikationen. *merz - medien + erziehung*, 61(3), S. 27–35.
- Rieger, D., Morten, A., & Frischlich, L. (2017). Verbreitung und Inszenierung. In L. Frischlich, D. Rieger, A. Morten, G. Bente, & in Kooperation mit der Forschungsstelle Terrorismus/Extremismus des Bundeskriminalamtes (Hrsg.), *Videos gegen Extremismus? Counter-Narrative auf dem Prüfstand* (S. 47–80). Wiesbaden: Bundeskriminalamt.
- Riesmeyer, C., Pfaff-Rüdiger, S., & Kümpel, A. S. (2016). Wenn Wissen zu Handeln wird: Medienkompetenz aus motivationaler Perspektive. *Medien & Kommunikationswissenschaft*, 64(1), S. 36–55.
- 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).

- 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, S. 461–470.
- Rost, K., Stahel, L., & Frey, B. S. (2016). Digital social norm enforcement: Online firestorms in social media. *PLOS ONE*, 11(6).
- Rowe, I. (2015a). Civility 2.0: A comparative analysis of incivility in online political discussion. *Information, Communication & Society*, 18(2), S. 121–138.
- Rowe, I. (2015b). Deliberation 2.0: Comparing the deliberative quality of online news user comments across platforms. *Journal of Broadcasting & Electronic Media*, 59(4), S. 539–555.
- Ruiz, C., Domingo, D., Micó, J. L., Díaz-Noci, J., Meso, K., & Masip, P. (2011). Public sphere 2.0? The democratic qualities of citizen debates in online newspapers. *The International Journal of Press/Politics*, 16(4), S. 463–487.
- S** Saleem, H. M., Dillon, K. P., Benesch, S., & Ruths, D. (2017, September 28). *A web of hate: Tackling hateful speech in online social spaces*. Abgerufen von <http://arxiv.org/abs/1709.10159> am 17.04.2019.
- Santana, A. D. (2014). Virtuous or vitriolic: The effect of anonymity on civility in online newspaper reader comment boards. *Journalism Practice*, 8(1), S. 18–33.
- Schenk, M. (2007). *Medienwirkungsforschung* (3. Auflage). Tübingen: Mohr Siebeck.
- Schmitt, J. B. (2017). Online-Hate Speech: Definition und Verbreitungsmotivationen aus psychologischer Perspektive. In K. Kaspar, L. Gräßer, & A. Riffi (Hrsg.), *Online Hate Speech. Perspektiven auf eine neue Form des Hasses* (S. 52–56). München: kopaed.
- Schmitt, J. B., Rieger, D., Rutkowski, O., & Ernst, J. (2018). Counter-messages as Prevention or Promotion of Extremism?! The Potential Role of YouTube. *Journal of Communication*, 68(4), S. 780–808.
- Schmitt, J. B., Riesmeyer, C., Ernst, J., Rieger, D., Nienierza, A., Fawzi, N., ... Roth, H.-J. (2019). Jugendliche und extremistische Propaganda. Kontakt - Schulische Prävention - Pädagogische Herausforderungen. *Praxis der Rechtspsychologie*, 29(1), S. 67–90.
- Schweiger, W. (2007). *Theorien der Mediennutzung. Eine Einführung*. Wiesbaden: Springer VS.
- Shachaf, P., & Hara, N. (2010). Beyond vandalism: Wikipedia trolls. *Journal of Information Science*, 36(3), S. 357–370.
- Sobieraj, S. (2018). Bitch, slut, skank, cunt: Patterned resistance to women's visibility in digital publics. *Information, Communication & Society*, 21(11), S. 1700–1714.
- Sood, S., Antin, J., & Churchill, E. (2012). Profanity use in online communities. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, S. 1481–1490.

- Spears, R., & Lea, M. (1994). Panacea or panopticon? The hidden power in computer-mediated communication. *Communication Research*, 21(4), S. 427–459.
- Sponholz, L. (2018). *Hate Speech in den Massenmedien*. Wiesbaden: Springer VS.
- Springer, N. (2014). *Beschmutzte Öffentlichkeit? Warum Menschen die Kommentarfunktion auf Online-Nachrichtenseiten als öffentliche Toilettenwand benutzen, warum Besucher ihre Hinterlassenschaften trotzdem lesen, und wie die Wände im Anschluss aussehen*. Berlin: Lit.
- Springer, N., & Kümpel, A. S. (2018). User-Generated (Dis)Content. Eine Literatursynopse zur Nutzung der Kommentarfunktion auf Nachrichtensites im Internet. In C. Nuernbergk & C. Neuberger (Hrsg.), *Journalismus im Internet. Profession – Partizipation – Technisierung* (2., aktualisierte und erweiterte Auflage, S. 241–271).
- Stieglitz, S., & Dang-Xuan, L. (2013). Emotions and information diffusion in social media—Sentiment of microblogs and sharing behavior. *Journal of Management Information Systems*, 29(4), S. 217–248.
- Stroud, N. J., Scacco, J. M., Muddiman, A., & Curry, A. L. (2015). Changing deliberative norms on news organizations' Facebook sites. *Journal of Computer-Mediated Communication*, 20(2), S. 188–203.
- Stryker, R., Conway, B. A., & Danielson, J. T. (2016). What is political incivility? *Communication Monographs*, 83(4), S. 535–556.
- Su, L. Y.-F., Xenos, M. A., Rose, K. M., Wirz, C., Scheufele, D. A., & Brossard, D. (2018). Uncivil and personal? Comparing patterns of incivility in comments on the Facebook pages of news outlets. *New Media & Society*, 20(10), S. 3678–3699.
- Suhay, E., Bello-Pardo, E., & Maurer, B. (2018). The polarizing effects of online partisan criticism: Evidence from two experiments. *The International Journal of Press/Politics*, 23(1), S. 95–115.
- Sukumaran, A., Vezich, S., McHugh, M., & Nass, C. (2011). Normative influences on thoughtful online participation. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, S. 3401–3410.
- Suler, J. (2004). The online disinhibition effect. *CyberPsychology & Behavior*, 7(3), S. 321–326.
- U** Upadhyay, S. R. (2010). Identity and impoliteness in computer-mediated reader responses. *Journal of Politeness Research. Language, Behaviour, Culture*, 6(1), S. 105–127.
- W** Waddell, T. F. (2018). What does the crowd think? How online comments and popularity metrics affect news credibility and issue importance. *New Media & Society*, 20(8), S. 3068–3083.

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*, S. 73–83.

Weber, P., & Wirth, W. (2013). Nachrichtenfaktoren und Relevanzattribution. Der Einfluss von Nachrichtenfaktoren auf Relevanzurteile von Rezipienten und die moderierende Rolle von Civic Pride. *Medien & Kommunikationswissenschaft, 61*(4), S. 514–531.

Wilson, J. Q., & Kelling, G. E. (1982). Broken windows. The police and neighborhood safety. *The Atlantic Monthly*.

Y Yavuz, C. (2017). Like - share - hate: Extremistische Umtriebe im Netz. *merz - medien + erziehung, 3*, S. 13–26.

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

Z Ziegele, M. (2016). *Nutzerkommentare als Anschlusskommunikation: Theorie und qualitative Analyse des Diskussionswerts von Online-Nachrichten*. Wiesbaden: Springer VS.

Ziegele, M., & Jost, P. B. (2016). Not funny? The effects of factual versus sarcastic journalistic responses to uncivil user comments. *Communication Research*. Advance Online Publication.

Zimmerman, A. G., & Ybarra, G. J. (2016). Online aggression: The influences of anonymity and social modeling. *Psychology of Popular Media Culture, 5*(2), S. 181–193.

Viele Diskussionen in sozialen Online-Medien sind durchzogen von Inzivilität; Hass und Beleidigungen scheinen mehr Regel als Ausnahme zu sein.

Die Ursachen für diese Wahrnehmung sind die einfache öffentliche Zugänglichkeit zu Online-Diskussionen, die Möglichkeit von schneller Verbreitung der Inhalte sowie fehlende regulative soziale Hinweisreize.

Nach bisherigen Studienbefunden sind die Wirkungen von Hassreden und inziviler Kommunikation in aller Regel negativ: Kommentierte Inhalte werden negativer beurteilt und es gibt verzerrte Vorstellungen über die Glaubwürdigkeit journalistischer Angebote und die Relevanz der Themen.

Präventive Strategien nehmen als Gegenmaßnahmen insbesondere langfristig die wichtigere Rolle gegenüber repressiver Maßnahmen ein. Aufgrund der großen Menge an Inhalten, werden dafür zur methodischen Analyse zunehmend *Maschine-Learning*-Techniken eingesetzt.