1.2 Der Einsatz von Moderation bei #meinfernsehen2021

Dominique Heinbach und Lena Wilms

1. Einleitung

Bürger:innenbeteiligungsverfahren werden als Schlüsselinstrument zur Konstituierung, Anhörung und Einbindung einer informierten Bürger:innenschaft angesehen (Newton, 2012; Frieß & Porten-Cheé, 2018). Herzstück solcher Beteiligungsverfahren im digitalen Raum sind dabei häufig Online-Diskussionen, in deren Rahmen Teilnehmende sich respektvoll und gleichberechtigt über Fragestellungen austauschen, Argumente abwägen und gemeinsam Lösungsansätze entwickeln sollen. Moderation wird dabei als adäquates Mittel gesehen, um solche qualitativ hochwertigen Debatten zu stimulieren (Blumler & Coleman, 2001; Wright & Street, 2007). Sie ist geeignet, um die deliberative Qualität der Debatte zu fördern, etwa indem sie Diskussionen zum Ausgangsthema zurückführt, rationalen Austausch anregt und Redeanteile verschiedener Interessengruppen ausbalanciert (Edwards, 2002; Epstein & Leshed, 2016; Wright, 2006; Smith, 2009). Gleichzeitig kann Moderation auch inziviles Verhalten, wie Beleidigungen oder Diskriminierungen, wirkungsvoll eindämmen, z. B. durch Löschen oder öffentliches Sanktionieren der entsprechenden Inhalte. So soll Moderation dazu beitragen, dass Teilnehmende einen geschützten Diskussionsrahmen für konstruktiven, offenen und niedrigschwelligen gegenseitigen Austausch vorfinden (Grimmelmann, 2015).

Auch bei #meinfernsehen2021 handelt es sich um ein diskursives Online-Beteiligungsverfahren, in dessen Rahmen Moderation zur Förderung der Diskussionsqualität eingesetzt wurde. Ein Team aus insgesamt sechs Moderator:innen bediente sich dabei einer breiten Palette an Moderationshandlungen, die über das Formulieren eigener Kommentare bis hin zur Löschung inadäquater Diskussionsinhalte reichte. Der vorliegende Beitrag widmet sich der systematischen Analyse dieser Moderation und nimmt sie gleichzeitig zum Anlass, eine extensive Systematik zur Beschreibung von Moderation in Beteiligungsverfahren vorzustellen. Die Analyse ist dabei entlang der Bestimmung sowohl der (1) Auswahlkriterien der moderierten Kommentare als auch (2) der zum Einsatz gekommenen Moderationsformen und -stile strukturiert. Wir knüpfen damit an kommunikati-

onswissenschaftliche Studien zur Identifikation von Moderationsfaktoren (Paasch-Colberg & Strippel, 2021) sowie zur Differenzierung verschiedener Formen und Stile von interaktiver und nicht-interaktiver Moderation (Wright, 2006; Ziegele et al., 2018) an. Mit dem Beitrag legen wir zudem die erste systematische Analyse von Moderation im Rahmen von diskursiven Online-Beteiligungsverfahren in Deutschland vor.

2. *Moderation von Online-Diskussionen*

Unter professioneller Moderation verstehen wir *Handlungen zur Strukturierung und Steuerung von Debatten zwecks Herbeiführung einer gewünschten Diskursatmosphäre durch die Diskursanbieter:innen* (in Anlehnung an Wright, 2006; Ziegele & Jost, 2020). Im Verständnis von Deliberationstheoretiker:innen ist eine gewünschte Diskursatmosphäre in der Regel eng mit der Vorstellung einer demokratischen Debatte verknüpft (Edwards, 2002; Wright, 2006). Entsprechend sollen die Strukturierung und Steuerung des Kommunikationsprozesses und dessen Inhalten auf das Ziel gerichtet sein, zu einem rationalen, reziproken und respektvollen Diskurs möglichst vieler unterschiedlicher Teilnehmer:innen anzuregen und inzivile Beiträge einzudämmen (Stroud et al., 2015; Wright, 2006; Ziegele & Jost, 2020). Moderator:innen werden dabei gängigerweise als Prozessbegleiter:innen (*engl.* facilitator; Grimmelmann, 2015) sowie als demokratische Intermediäre (Edwards, 2002) charakterisiert, die im Rahmen ihrer Tätigkeit unterschiedliche Funktionen erfüllen: In ihrer *strategischen Funktion* definieren sie den Diskursrahmen mit Blick auf das Ziel der Diskursanbieter:innen, indem sie etwa Ziele und Themen der Debatte festlegen und die Ergebnisse der Diskussion an die Anbieter:innen rückbinden. Im Rahmen der *Bereitstellungsfunktion* (*engl.* conditioning function) stellen sie zusätzliche Informationen bereit und stimulieren Partizipation der Teilnehmenden im Vorfeld der Diskussion. Alle Tätigkeiten zur Betreuung der laufenden Diskussion, etwa durch Formulieren von Diskussionsregeln und deren Durchsetzung, werden als *Prozessfunktion* bezeichnet (Edwards, 2002).

Moderation wird in der Praxis vornehmlich im Sinne der Prozessfunktion verstanden. In diesem Kontext dient Moderation als Sammelbegriff für eine Vielzahl von Systemen und Techniken, die von Anbieter:innen in Online-Foren zur Anwendung gebracht werden können. Demgegenüber steht ein überschaubarer Forschungsstand, der sich mit der empirischen Abbildung dieser vielfältigen Moderationstechniken im Kontext von deliberativen Online-Beteiligungsverfahren beschäftigt. Hier wird Moderation gängigerweise als Designmerkmal der Diskussionsplattform erfasst (Frieß

& Eilders, 2015). Die Analyse von Moderation beschränkt sich dabei meist auf ihr Vorhandensein bzw. die Differenzierung verschiedener Moderationsformen (z. B. interaktiv vs. nicht-interaktiv, s.u.; Epstein & Leshed, 2016; Frieß, 2016; Wright, 2006). Eine Systematisierung zur Analyse verschiedener Moderationsstrategien jenseits dieser Differenzierung fehlt allerdings weitestgehend. Das vorliegende Forschungsvorhaben zielt darauf ab, die Analyse von Moderation systematisch zu erweitern. Wir schlagen vor, Moderation als zweistufigen Prozess zu verstehen, der (1) aus der systematischen Auswahl der zu moderierenden Kommentare sowie (2) aus der Wahl der geeigneten moderativen Handlung besteht (siehe hierfür auch Einwiller & Kim, 2020; Paasch-Colberg & Strippel, 2021). Eine systematische Beschreibung von Moderation soll entsprechend entlang dieser beiden Stufen erfolgen.

2.1. Stufe 1: Auswahlkriterien für Moderation

Die Auswahl der zu moderierenden Kommentare ist als erste Stufe der Moderation zu begreifen. Fraglich ist, nach welchen Kriterien die Auswahl moderierter Kommentare erfolgt und ob diese Auswahl konsistent und exhaustiv angewandt wird. Die Erforschung solcher Auswahlkriterien für Moderation ist aktuell noch recht jung und beschränkt sich ausschließlich auf die Moderation von inzivilen Kommentaren. In einer Inhaltsanalyse der Kommentarspalten von Online-Nachrichtenseiten von Ziegele et al. (2018) erhöhte die Inzivilität in Initialkommentaren die Wahrscheinlichkeit, dass unter den Antwortkommentaren mindestens ein Moderationskommentar war.[1] Das galt jedoch nur für „public level incivility", die negative Stereotype, Lügenvorwürfe und Gewaltandrohungen enthielt. „Personal level incivility" in Form von vulgärer Sprache, Beleidigungen, Sarkasmus und „Schreien" hatte keinen Einfluss auf die Moderationsentscheidung. Stroud und Muddiman (2017) konzentrieren sich auf die Analyse von gelöschten Kommentaren im Online-Diskurs der *New York Times*. Im Zentrum steht die Fragestellung, welche Kommentarmerkmale die Moderationswahrscheinlichkeit erhöhen. Dort konnte nachgewiesen werden, dass ein signifikanter Zusammenhang zwischen dem Vorkommen bestimmter Schimpfworte und der Löschung von Kommentaren besteht. Boberg et al. (2018) können einen solchen Zusammenhang im Rahmen

1 Initial- und Antwortkommentar können synonym zu den Begrifflichkeiten Top-Level Kommentar und Sub-Level Kommentar verwendet werden.

einer Analyse der Kommentare des deutschsprachigen Online-Diskurses von *Spiegel Online* nicht nachweisen, finden jedoch Hinweise darauf, dass Moderation von Inzivilität deutlich restriktiver entlang spezifischer Themen (Flüchtlingskrise, Fake News und Rechtspopulismus) eingesetzt wird. In beiden Studien wird eine mangelnde Konsistenz in der Anwendung von Moderation konstatiert, die (teilweise) auch in der Studie von Ziegele et al. (2018) deutlich wird (Boberg et al., 2018; Stroud & Muddiman, 2017). Mögliche Gründe für diese inkonsistenten Moderationshandlungen liefern Paasch-Colberg und Strippel (2021). Diese führen Erkenntnisse aus 23 qualitativen Expert:inneninterviews mit Community-Manager:innen verschiedener journalistischer Online-Angebote zusammen und identifizieren auf diesem Wege Faktoren, die den Auswahlprozess der zu moderierenden Kommentare beeinflussen können. Neben erlernten Regeln entlang professioneller Standards und Routinen sowie rechtlichen Rahmenbedingungen schlagen sich auch individuelle Faktoren, wie z. B. Persönlichkeitsmerkmale und Organisationsrichtlinien, in der Kommentarauswahl nieder. Einwiller und Kim (2020) weisen darauf hin, dass sich Organisationsrichtlinien (Netiquetten) zur Löschung von Hate Speech häufig auf die Aufzählung von unzulässigen Kommunikationsformen (z. B. Beleidigungen, Vulgarität, Gewaltandrohungen) beschränken. Die Identifikation dieser Merkmale in Kommentaren liegt jedoch im Ermessen der Moderator:innen.

Auch im Rahmen des #meinfernsehen 2021 Projektes soll untersucht werden, welche Kommentare moderiert wurden. Es soll ermittelt werden, ob sich moderierte Kommentare systematisch von nicht moderierten Kommentaren unterscheiden, um so Rückschlüsse auf Systematik und Konsistenz der Moderation anstellen zu können. Im Gegensatz zum Forschungsstand beschränken wir uns dabei jedoch nicht nur auf Hasskommentare, sondern beziehen ausdrücklich alle Kommentare mit ein, die eine Moderation erhalten haben. So können auch Kriterien jenseits von Inzivilität identifiziert werden, die einen Einfluss auf die Moderationswahrscheinlichkeit ausgeübt haben.

FF1: *Inwiefern unterscheiden sich Kommentare, die moderiert wurden, von nicht moderierten Kommentaren?*

2.2. Stufe 2: Formen und Stile von Moderation

Moderationsformen und -stile beschreiben, wie eine moderative Handlung konkret ausgestaltet ist. In Anlehnung an Wright (2006) können dabei folgende Moderationsformen unterschieden werden: Bei *nicht-interaktiver*

Moderation handelt es sich um eine unsichtbare Moderationsform, bei der unzulässige Beiträge, die bspw. gegen die Netiquette verstoßen, teilweise oder gänzlich entfernt werden (Wright, 2006). Dies trifft üblicherweise auf Kommentare mit inzivilen, also beispielsweise beleidigenden oder diskriminierenden, Inhalten zu. Nicht-interaktive Moderation verzichtet auf die aktive Teilnahme an Diskussionen. Bei *interaktiver Moderation* beteiligen sich die Moderierenden hingegen aktiv mit eigenen Beiträgen an der Diskussion (Stroud et al., 2015; Ziegele & Jost, 2020).

Den Moderator:innen steht dabei eine breite Palette an Interaktionen zur Verfügung, die je nach Moderationsanlass und -ziel variieren können. Im Rahmen der interaktiven Moderation können zwei übergeordnete Stile unterschieden werden: *Regulierende Moderation* verfolgt das Ziel, die Einhaltung der Netiquette durchzusetzen und gegenseitige Normen des Respekts zu etablieren. Im Rahmen der Kommentare wird dabei öffentlich auf eine Verletzung der Netiquette hingewiesen. Ferner kann die *regulierende Moderation* auch Diskussionen zum Ausgangsthema zurückführen und auf Kritik und Fragen an die Redaktion reagieren (Ziegele & Jost, 2020). Bei der *unterstützenden Moderation* wird der Fokus zusätzlich auf deliberative Kommentare, d.h. rationale und respektvolle Bezugnahmen (Gutmann & Thompson, 1996), gelegt. Moderator:innen nehmen hier aktiv an Debatten teil, indem sie etwa Fragen der Nutzer:innen beantworten, Zusatzinformationen in den Diskurs einbringen, deliberative Kommentare lobend hervorheben und mit Anschlussfragen eine sachliche Debatte vorantreiben (Stroud et al., 2015; Wright, 2006; Ziegele & Jost, 2020). Sowohl unterstützende als auch regulierende Moderation bezeichnen Ziegele et al. (2018) dann als deliberativ, wenn moderative Handlungen auf die Einhaltung deliberativer Standards gerichtet sind (z. B. Eindämmung von Inzivilität) oder sich die Einhaltung von Rationalitäts-, Reziprozitäts- und Zivilitätsnormen im eigenen Kommentierverhalten widerspiegelt.

Kommunikationswissenschaftliche Studien weisen darauf hin, dass die Wirkung von Moderation auf das Kommentierverhalten der Nutzer:innen auch vom Moderationsstil abhängt. So zeigt die inhaltsanalytische Studie von Ziegele et al. (2018; 2019), dass ein unterstützender Moderationsstil, welcher sich durch das Einbringen von zusätzlichen Informationen, inhaltlichen Argumenten und Fragen auszeichnet, besonders für die Steigerung von Rationalität in der Diskussion geeignet war. Hiervon abzugrenzen ist informelle Moderation mit Smalltalk, Humor und Lob („sociable"), die von den Autor:innen als nicht-deliberativ bezeichnet wird und keinen Effekt auf die Rationalität der Folgediskussion hatte. Regulierende oder sogar konfrontative Moderation, die auf die Bloßstellung von Teilnehmenden abzielt, könne indes zu einer Steigerung von Inzivilität in den Fol-

gekommentaren führen (Ziegele et al., 2018; Ziegele et al., 2019). Die Anwendung verschiedener Moderationsstile hatte zudem einen Effekt auf Wahrnehmungen und Einstellungen der Nutzer:innen: Experimentalstudien zeigten, dass sich faktenorientierte und empathische Moderation positiv auf das subjektive Gruppenzugehörigkeitsgefühl zur Community und die Bewertung des Diskursanbieters auswirken kann, während sarkastische Moderation negative Wirkungen auf die genannten Wahrnehmungen entfaltete (Masullo et al., 2021; Ziegele & Jost, 2020).

In der systematischen Analyse sollen die moderativen Handlungen im Rahmen des #meinfernsehen2021-Projektes inhaltlich untersucht werden und, wenn möglich, in bestehende Systematiken zu Moderationsformen und -stilen eingeordnet werden.

FF2: *Welche Formen und Stile von interaktiver Moderation kamen bei #meinfernsehen2021 zur Anwendung?*

3. Untersuchungsgegenstand: Moderation bei #meinfernsehen2021

Im Rahmen des #meinfernsehen2021-Projektes wurde sowohl interaktiv als auch nicht-interaktiv moderiert. Für die vorliegende Studie sind dabei diejenigen moderativen Handlungen von Interesse, die unmittelbar zur Betreuung der laufenden Debatte durchgeführt wurden (Prozessfunktion, Edwards, 2002). Im Verfahrenszeitraum waren insgesamt sechs professionelle Moderator:innen beschäftigt, dabei maximal vier zeitgleich. Die Betreuung der Plattform erfolgte wochentags von 8 bis 20 Uhr durchgängig, am Wochenende bei Bedarf. Die Moderator:innen wurden dabei via E-Mail über neu eingehende Kommentare informiert. Zusätzlich stand dem Moderationsteam ein gemeinsamer Chat-Kanal zur Verfügung, um sich teamintern zu koordinieren und auszutauschen.

Die Moderation erfolgte dabei entlang vorab formulierter Leitlinien. Diese definierten folgende Ziele: (1) Die Initiierung von Diskussionen und Motivation der Diskutierenden sowie (2) die Einhaltung der für die Plattform gültigen Diskussionsregeln (Netiquette) und der rechtlichen Bestimmungen. Ersteres sollte über einen respektvollen, wertschätzenden und konstruktiven Kommunikationsstil erreicht werden. Bei Letzterem galt es, kontroverse Standpunkte im Sinne des freien Meinungsaustausches zuzulassen, solange sie nicht gegen die AGB der Plattform oder die Netiquette verstoßen und es sich nicht um strafrechtlich relevante Inhalte (Beleidigung, verfassungswidrige Kennzeichen, Volksverhetzung) handelte. Auswahlkriterien zur Identifikation von moderationswürdigen Kommentaren sind nicht explizit definiert. Implizit wird aber das Löschen von straf-

rechtlich relevanten Inhalten sowie das Sanktionieren diskriminierender und beleidigender Kommentare vorgeschrieben.

4. Methode

Insgesamt wurden im Partizipationsverfahren 3.817 Kommentare verfasst. 384 Nutzer:innenkommentare wurden moderiert (9.5 %). Dabei wurde in 23 Fällen (0.6 %) nicht-interaktive Moderation eingesetzt, indem die Kommentare von der Moderation gelöscht wurden. 361 Kommentare wurden interaktiv moderiert, indem die Moderator:innen mit einem oder mehreren eigenen Kommentaren auf den jeweiligen Teilnehmer:innenkommentar reagierten. Insgesamt lagen 373 Moderationskommentare vor. Das entspricht einem Anteil von 9.8 Prozent an der Gesamtdiskussion.

Um FF1 zu beantworten, wurde eine quantitative Inhaltsanalyse von moderierten und nicht moderierten Kommentaren durchgeführt. Dafür wurden zusätzlich zu der geschichteten Zufallsstichprobe (siehe Methodenkapitel in diesem Sammelband) alle Kommentare, die moderiert wurden, in die Stichprobe aufgenommen. Bedauerlicherweise wurden gelöschte Kommentare nicht archiviert, sodass nicht-interaktiv moderierte Kommentare nicht inhaltsanalytisch untersucht werden konnten. Deshalb wurden in der quantitativen Inhaltsanalyse nur interaktiv moderierte Kommentare berücksichtigt. Dieses Vorgehen resultierte in einer Gesamtstichprobe von 1.682 Kommentaren. Für die inferenzstatistischen Analysen wurden die zentralen Kategorien aus dem Codebuch (siehe Methodenkapitel in diesem Sammelband) zu Mittelwert-Indizes zusammengefasst, die die unterschiedlichen Dimensionen von deliberativer Qualität sowie Inzivilität abbilden sollen (Coe et al., 2014; Esau et al., 2019; Frieß & Eilders, 2015; Ziegele et al., 2020): *Rationalität* (Themenbezug; Tatsachenbehauptung; Begründung; Lösungsvorschlag; Zusatzwissen; Frage; M = 1.9, SD = 0.5), *Reziprozität* (Bezugnahme Nutzer:in/Community; Bezugnahme Inhalt; M = 2.0, SD = 1.3), *expliziter Respekt* (höfliche Anrede; Respektsbekundungen; M = 1.2, SD = 0.4), *Persönliche Erfahrungen/Storytelling* (Einzelkategorie, kein Index; M = 1.1, SD = 0.5), *Emotionen* (positive und negative Emotionen; M = 1.4, SD = 0.6) und *Inzivilität* (Geringschätzung; Schreien;

Vulgarität; Beleidigungen; Sarkasmus/Zynismus/Spott; Lügenvorwürfe; M = 1.2, SD = 0.4).[2]

Um FF2 zu beantworten, wurden alle Moderationskommentare (n = 373), die im Laufe des Partizipationsverfahrens geschrieben wurden, mit einer inhaltlich-strukturierenden qualitativen Inhaltsanalyse (Kuckartz, 2018) ausgewertet. Das Material wurde computergestützt mit dem Programm „MaxQDA" codiert. Dabei wurden zunächst Kategorien deduktiv aus den theoretischen Überlegungen, dem Forschungsstand und den Forschungsfragen abgeleitet (z. B. regulierende und unterstützende Moderation, Fragen, Zusatzwissen, Lob). Diese wurden anschließend induktiv am Datenmaterial weiterentwickelt, ausdifferenziert und um weitere Kategorien ergänzt.

5. Ergebnisse

5.1. Auswahl der moderierten Kommentare

Um zu untersuchen, ob und inwiefern sich interaktiv moderierte Kommentare hinsichtlich ihrer deliberativen Qualität und Inzivilität von nicht moderierten Kommentaren unterscheiden, wurde eine logistische Regressionsanalyse mit den Qualitätsdimensionen als unabhängige Variablen und der Moderationsentscheidung (moderiert/nicht moderiert) als abhängige Variable durchgeführt. Wie bereits erwähnt gingen nur interaktiv moderierte Kommentare in die Analyse ein. In Tabelle 1 ist zu sehen, dass Inzivilität der stärkste Prädiktor für Moderation war (b = 0.61, Odds = 1.85, p <.001). Außerdem zeigte sich ein positiver Effekt von Rationalität auf die Moderationsentscheidung (b = 0.43, Odds = 1.53, p <.001). Letztere entsprechen damit eher deliberativen Qualitätsanforderungen mit Blick auf eine rationale Diskussion. Reziprozität, expliziter Respekt, Emotionen und Storytelling hatten keinen Einfluss auf die Moderationsentscheidung.

2 Die Erfassung der einzelnen Merkmale erfolgte auf einer Skala von 1 = „Merkmal ist eindeutig nicht vorhanden" bis 4 = „Merkmal ist eindeutig vorhanden" (vgl. Methodenkapitel in diesem Sammelband).

Tabelle 1. *Logistische Regression von Qualität und Inzivilität der Kommentare auf die Moderationsentscheidung*

Abhängige Variable: Kommentar wurde moderiert	b	95% CI für Odds Ratio		
		Unterer Wert	Odds	Oberer Wert
Konstante	-3.38 *** [-4.14, -2.66]			
Rationalität	**0.43** *** [0.20, 0.68]	1.21	1.53	1.94
Reziprozität	0.07 [-0.04, 0.17]	0.97	1.07	1.18
Expliziter Respekt	0.17 [-0.17, 0.45]	0.90	1.19	1.59
Emotionen	-0.06 [-0.29, 0.14]	0.76	0.94	1.17
Storytelling	0.20 [-0.06, 0.42]	0.96	1.22	1.54
Inzivilität	**0.61** *** [0.30, 0.92]	1.35	1.85	2.54
n		1663		

*R^2 =.02 (Cox & Snell).03 (Nagelkerke). Model χ^2 = 35.93, p <.001. * p < 0.05; ** p < 0.01; *** p < 0.001.*

Die Inhalte der gelöschten Kommentare standen bedauerlicherweise nicht für die Inhaltsanalyse zur Verfügung, die Gründe für die Löschung lassen sich aber zum Großteil auf Basis der qualitativen Inhaltsanalyse der Moderationskommentare rekonstruieren, da die Moderator:innen die Löschungen in fast allen Fällen in der Diskussion transparent machten. Nur in einem Fall war die Moderationsentscheidung nicht nachvollziehbar, da auch der Moderationskommentar gelöscht wurde. Der häufigste Grund für eine Löschung war Spamming, also das wiederholte Posten inhaltlich identischer Kommentare. Das ist auch eine mögliche Erklärung dafür, warum 18 der 23 gelöschten Kommentare von einem Account stammten: Die betreffende Person scheint aufgrund eines Missverständnisses denselben Text 21-mal gepostet zu haben. Lediglich ein Kommentar wurde wegen Beleidigungen gelöscht. Dieser Kommentar liegt auch als Screenshot vor, der von den Moderator:innen archiviert wurde.[3] Der betreffende Nutzer hatte eine andere Person als „dumme Nuss" bezeichnet. Zudem wurde in zwei Fällen ein Kommentar durch die Moderation bearbeitet. In einem Fall wurde ein laut Netiquette nicht zulässiger Link entfernt (fehlendes Impressum), im anderen ein Buchcover aufgrund möglicher Urheberrechte.

3 Die Moderator:innen hatten in ihrem gemeinsamen Chat-Kanal über den Umgang mit diesem Kommentar diskutiert.

Die Ergebnisse zeigen, dass nicht nur inzivile Kommentare, die gegen die Netiquette verstoßen hatten, moderiert wurden, sondern auch Kommentare, die eine vergleichsweise hohe Rationalität und damit eine zentrale Dimension deliberativer Qualität aufwiesen. Wie diese Moderation konkret aussah, sollen die Ergebnisse der qualitativen Inhaltsanalyse der Moderationskommentare zeigen.

5.2. Formen und Stile interaktiver Moderation

Zur Beantwortung der zweiten Forschungsfrage wurden alle 373 Moderationskommentare qualitativ untersucht. Eines der am häufigsten auftretenden Moderationsmerkmale waren *Fragen*, die in 182 Moderationskommentaren vorkamen (48.8 %). Dabei stellten die Moderator:innen vor allem *Nachfragen nach Belegen, Beispielen, Konkretisierungen oder Begründungen* (67 Fälle bzw. 18 %, „Was genau meinen Sie mit ‚schlechter Nutzbarkeit'? Können Sie dies konkretisieren?", MK 1545[4]). In 43 Fällen (11.5 %) fragten sie die Teilnehmer:nnen aber auch nach ihren *persönlichen Meinungen oder Wünschen* („Wie oft pro Woche sollten diese Dokus Ihrer Meinung nach ausgestrahlt werden?", MK 1744). Zudem baten sie in 24 Moderationskommentaren (6.4 %) um *(Lösungs-) Vorschläge* („Wie könnte eine Reform in Deutschland aussehen? Was wären für Sie die wichtigsten Punkte?", MK 1301). In einigen Fällen wurden Fragen auch genutzt, um die Aussagen oder Standpunkte der Nutzer:innen zu *konkretisieren oder zusammenzufassen* ("[...] dass [sic!] heißt, Sie sprechen sich gegen kostenpflichtige Funktionen – wie Werbung – aber nicht gegen die kostenfreie Nutzung von Online-Plattformen aus?", MK 2103). Sehr selten wurden *Verständnisfragen* gestellt (acht Fälle bzw. 2.1 %, „Verstehe ich Sie richtig, dass Sie die Inhalte der Mediatheken bevorzugt auf dem Handy konsumieren?", MK 1164) und nach persönlichen Erfahrungen gefragt (drei Fälle bzw. 0.8 %, „Nutzen Sie neben Videotext auch andere Zusatzinformationen, die die Sender über das Fernsehgerät anbieten – etwa HbbTV?", MK 1423). Insgesamt entstand der Eindruck, als ob die Fragen vor allem das Ziel verfolgten, die Diskussion voranzubringen und beispielsweise zusätzliche Argumente, Meinungen, Informationen und Vorschläge zu fördern. In einigen Fällen erfüllten Fragen aber auch eine *rhetorische Funktion*, mit der unterschwelli-

4 Die Belege in der qualitativen Analyse setzen sich aus einem Kürzel für die Art des Kommentars (MK = Moderationskommentar, TK = Teilnehmer:innenkommentar) und der eindeutigen Kommentar-ID zusammen.

ge Kritik geäußert und subtile Gegenrede betrieben wurde, ohne zu offensiv „dagegenzuhalten" („Was genau meinen Sie mit ‚trivialem Dummfunk'?", MK 1596).

112 Moderationskommentare (30 %) enthielten *Zusatzinformationen*. Es wurden z. B. *Studien oder Artikel zu den Diskussionsthemen verlinkt* („Für alle, die den Beitrag gern nachlesen möchten, füge ich den Link ein: [...]", MK 1914), *zusätzliche Fakten geliefert* („[...] der öffentlich-rechtliche Rundfunk ist bislang staatsvertraglich verpflichtet, Unterhaltungsprogramme anzubieten", MK 2234) und *Fragen der Teilnehmer:innen beantwortet* („Hallo [Nutzer:in], beim Sprechen wird wie beim Gendersternchen eine kleine Pause zwischen Wortstamm und Endung gelassen", MK 1772). Zudem wurde in 17 Fällen (4.6 %) *auf andere Threads verwiesen*, um die Diskussionen zu strukturieren („Zum Thema Vertonung und Hörprobleme gibt es hier eine Diskussion, [Link], die für Sie vielleicht auch interessant wäre", MK 1272).

Transparenz zum Partizipationsverfahren und der Moderation spielte in 37 Moderationskommentaren (9.9 %) eine Rolle. Es wurde beispielsweise erläutert, wer das Verfahren ausrichtet und wann und wo die Ergebnisse vorgestellt werden. („Diese Diskussionsplattform wird nicht von den öffentlich-rechtlichen Sendern betrieben, sondern ist ein Projekt des Grimme-Instituts in Kooperation mit der Bundeszentrale für politische Bildung und dem Düsseldorfer Institut für Internet und Demokratie [...]", MK 1808). Zudem wurde in einigen Fällen *auf Kritik eingegangen* („Ich habe Ihre Kritik entsprechend weitergeleitet, wir melden uns schnellstmöglich!", MK 2964) und *Moderationsentscheidungen begründet* („Ich habe den Link aufgrund der Netiquette [...] entfernt, weil nicht klar ist, ob es sich um ein legales Angebot handelt und das Impressum unvollständig ist", MK 1384).

29 Moderationskommentare (7.8 %) verfolgten erkennbar die Intention, *Falschinformationen oder verzerrte Informationen zu korrigieren, richtigzustellen oder ins Verhältnis zu setzen.* („Hallo [Nutzer:in], wir hegen Zweifel an Ihrer Aussage, dass in den öffentlich-rechtlichen [sic!] Nachrichten von ‚nur' 86 Cent gesprochen wird bzw. wurde. Haben Sie eventuell eine Quelle, die Ihre Aussage bestätigen würde?", MK 1594). Dabei wurde den Teilnehmer:innen aber in der Regel nicht vorgeworfen, absichtlich Desinformation zu betreiben, sondern in der Formulierung davon ausgegangen, dass es sich um nicht intendierte Fehler oder Verzerrungen handelte oder dass die Person sich ungenau ausgedrückt hat. („Nur zur Klarstellung: Der öffentlich-rechtliche Rundfunk [...] wird (mit Ausnahme der Deutschen Welle) nicht über Steuern finanziert, sondern über einen Rundfunkbeitrag", MK 442).

Die meisten Moderationskommentare waren sehr höflich und wertschätzend formuliert. In 196 Fällen (52.6 %) *bedankten* sich die Moderator:innen bei den Teilnehmer:innen für ihre Beiträge („danke für Ihre Beiträge auf dieser Plattform!", MK 952). 232 Moderationskommentare (62.2 %) enthielten *Begrüßungen, Verabschiedungen oder Höflichkeitsfloskeln* („Wünsche einen schönen Tag und eine gewinnbringende Diskussion", MK 1071). Außerdem wurden in 261 Fällen (70 %) die betreffende Person *direkt mit dem Nutzer:innennamen angesprochen* („Hallo [NutzerIn], haben Sie vielen Dank für Ihren Beitrag!", MK 1247). 50 Moderationskommentare (13.4 %) enthielten *Lob* („[...] danke für Ihren wertvollen Beitrag!", MK 956; „Ihr Vergleich zu einer Art ‚Escape-Room' klingt sehr interessant!", MK 1247). *Explizite Zustimmung* kam allerdings sehr selten vor (acht Fälle bzw. 2.1 %, „Sie haben absolut recht: Auch auf dieser Diskussionsplattform ist deutlich zu erkennen, dass die Teilnehmer*innen sehr dankbar für die Möglichkeit sind, sich untereinander auszutauschen, Ideen einzubringen und somit ‚gehört' zu werden", MK 2568), ebenso wie *Empathie* (drei Fälle bzw. 0.8 %, „Dass man die Fragestellungen, so wie sie [sic!] diese dargelegt haben, missverständlich auffassen könnte, verstehen wir", MK 1734).

In 40 Moderationskommentaren (10.7 %) wurde nicht nur auf eine einzelne Person eingegangen, sondern versucht, *weitere Teilnehmer:innen anzusprechen* und in die Diskussion einzubeziehen. Die Ansprache mehrerer Teilnehmer:innen wurde häufig genutzt, um mehrere Kommentare gebündelt zu moderieren („Liebe Diskussionsteilnehmer*innen, haben Sie vielen Dank für den regen Austausch. Ich möchte Sie darum bitten, auf einen respektvollen und wertschätzenden Ton zu achten [...]" [MK 3085]. Zudem wurde in einigen Moderationskommentaren das „*Gemeinwohl*" angesprochen und Vorteile von Informationen und Verhaltensweisen für die gesamte #meinfernsehen2021-Community herausgestellt („[...] belegen Sie Behauptungen möglichst mit Argumenten und Fakten. Ziehen Sie Quellen heran. So entsteht eine produktivere Diskussion. Davon profitieren alle", MK 1072; „hier noch der Link zum neuen Vorschlag von Ihnen, damit alle diesen finden", MK 830; „[...] gibt es einen Link dazu? Das könnte sicherlich für viele Diskussionsteilnehmer:innen interessant sein nachzulesen", MK 3231). Vereinzelt wurde auch versucht, *weitere Teilnehmer:innen aktiv zur Beteiligung zu motivieren* („Wir sind gespannt auf weitere Meinungen zu diesem Thema!", MK 1177).

67 Moderationskommentare (18 %) enthielten *Hinweise auf Verstöße gegen geltende Diskursnormen*. Darunter fielen vor allem *Reaktionen auf unhöfliches Verhalten* wie Beleidigungen, ein respektloser Ton und mangelnde Wertschätzung (26 Fälle bzw. 7 %, „Ich möchte Sie bitten, auf einen freundlichen und wertschätzenden Ton zu achten", MK 2656), und *Themenabwei-*

chungen (24 Fälle bzw. 6.4 %, „Ihr Beitrag führt leider an der Fragestellung vorbei", MK 2947). In 10 Fällen (2.7 %) wurde auf unsachliche und polemische Aussagen, fehlende Argumente und Belege hingewiesen („[...] bitte achten Sie auf einen sachlichen, wertschätzenden Ton und erläutern Sie Ihre Argumente nachvollziehbar", MK 1710). Seltener enthielten Moderationskommentare Hinweise auf *Spamming* (sechs Fälle bzw. 1.6 %, „Dennoch muss ich Sie darauf hinweisen, dass ‚Copy+Paste' in einem Diskussionsforum oftmals wenig zielführend ist", MK 2659), und *Falschinformationen* (vier Fälle bzw. 1.1 %, „Dort steht, dass Gottschalk in den 80ern festangestellt war. Von ‚ein paar Monaten', wie Sie sagen, ist dort nichts zu lesen", MK 2959). In zwei Fällen (0.5 %) gab es auch Hinweise auf *Holocaust-Relativierungen und Nazi-Vergleiche* („Ihre Aussage, dass die ‚Nachrichtensendungen Tagesschau, heute und so weiter [...] zuviel Propaganda' bringen ‚wie eben Nazipropaganda', ist höchst missverständlich und wie ich finde eine unzulässige Gleichsetzung", MK 641). Insbesondere auf Verstöße gegen Höflichkeitsnormen und Holocaust-Relativierungen reagierten die Moderator:innen häufig mit *regulierender Moderation* (46 Fälle bzw. 12.3 %) wie beispielsweise *Ermahnungen* („Bitte achten Sie auch einen sachlichen Diskussionsstil und vermeiden Sie Beleidigungen (‚Spinner')", MK 1072) und deutlicher inhaltlicher *Gegenrede* („Die Gleichsetzung der verbrecherischen NS-Ideologie, die zu einem Völkermord an sechs Mio. Menschen geführt hat, mit einer geschlechtergerechten Sprache halte ich für eine gefährliche Verharmlosung", MK 247). In elf Fällen (3 %) wurde *auf die Netiquette und die geltenden Diskursregeln verwiesen* („Hallo [Nutzer:in], beachten Sie bitte die Netiquette [Link] und verzichten Sie auf Schimpfwörter – danke", MK 1378). Lediglich in einem Fall wurden Sanktionen angedroht („[...] bitte achten Sie hier auf faire Umgangsformen und verzichten Sie auf pauschale Urteile zu anderen Diskussionsteilnehmer:innen [...]. Sonst werden solche Beiträge entsprechend der Netiquette gelöscht [Link]", MK 3484). 70 Prozent der überwiegend regulierenden Moderationskommentare enthielten allerdings trotzdem *unterstützende Elemente* wie Danksagungen, Höflichkeitsfloskeln oder die Bestärkung unterstützenswerter Aspekte (32 Fälle, „Nachfragen sind immer willkommen und tragen ja zu einer guten Diskussion bei. Dafür ist es aber nicht erforderliche, die Meinung anderer in unsachlicher Weise zu benennen. Vielen Dank und weiterhin gute Diskussionen!", MK 1176). Es ist allerdings anzumer-

ken, dass Inzivilität insgesamt nur selten vorkam.[5] Insgesamt entstand der Eindruck, dass die Moderation bei Normverstößen sehr streng war und auch bei vergleichsweise milden Verstößen oder „Grenzfällen" schnell eingriff. Einige Nutzer:innen äußerten auch die Kritik, dass sie die Moderation teilweise nicht nachvollziehen konnten oder ungerechtfertigt fanden („[...] der werte [Nutzer] hat somit nur Formulierungen aufgegriffen, welche im OT bereits benutzt – und bis dato nicht moniert worden sind [sic!] – Bitte machen Sie also mit aller Höflichkeit [sic!] und Respekt Ihren Moderatorenjob vernünftig …", TK 1936).

6. Diskussion

Insgesamt war die Moderation im Partizipationsverfahren #meinfernsehen2021 mit einem Anteil von fast zehn Prozent sehr präsent und aktiv. Bezüglich der Auswahl der moderierten Kommentare (FF1) zeigen die Ergebnisse der quantitativen Inhaltsanalyse, dass sowohl inzivile Kommentare moderiert wurden als auch Kommentare, die eine vergleichsweise hohe Rationalität aufwiesen und damit eine zentrale Voraussetzung für einen deliberativen Diskurs erfüllten (Frieß & Eilders, 2015). Dementsprechend ist das Ergebnis der qualitativen Analyse, dass überwiegend in einem unterstützenden Stil moderiert wurde (FF2), vor dem Hintergrund des insgesamt geringen Inzivilitätsanteils in den Diskussionen wenig überraschend. Es wurden überwiegend Formen von Moderation eingesetzt, die Ziegele et al. (2018) als „deliberativ" einordnen: Nach der Taxonomie der Autor:innen wurde hauptsächlich „diskursive" und „regulative" Moderation eingesetzt. „Konfrontative" (z. B. sarkastische) Moderation kam überhaupt nicht zum Einsatz. Aspekte einer „sociable" Moderation waren nur in Form von Lob und einem höflichen Umgangston zu erkennen. Die Autor:innen ordnen diese Form der Moderation als „nicht-deliberativ" ein, da sie nicht auf Rationalität abzielt. Ein höflicher Umgangston, Lob und Wertschätzungen können allerdings einen respektvollen Umgang der Teilnehmer:innen untereinander fördern, was wiederum als Voraussetzung für Deliberation gilt (Habermas, 1983; Steiner, 2012; Ziegele et al., 2020).

Wir argumentieren, dass unterschiedliche Aspekte von Moderation das Potenzial haben, unterschiedliche Dimensionen von deliberativer Qualität

5 18.7 % der Kommentare enthielten Geringschätzungen („Merkmal ist eindeutig vorhanden" und „Merkmal ist eher vorhanden"), 5.3 % Schreien, 1.9 % vulgäre Sprache, 2.3 % Beleidigungen und 14.2 % Sarkasmus, Zynismus oder Spott.

zu fördern. Die meisten Moderationskommentare im Verfahren zielten auf die Förderung von *Rationalität* ab: Es wurde z. B. nach Argumenten, Quellen und Beispielen gefragt, Zusatzinformationen geliefert, zusätzliche Aspekte in die Diskussion eingebracht und falsche, verzerrte oder unvollständige Informationen korrigiert. Außerdem wurde darauf geachtet, dass die Diskussionen beim Thema des jeweiligen Posts blieben. Das geschah überwiegend in einem unterstützenden Ton. Nur vereinzelt wurden Teilnehmer:innen ermahnt, beim Thema zu bleiben oder ihre Aussagen mit Argumenten zu belegen. Zudem zielten Moderationskommentare darauf ab, *Inzivilität* einzudämmen. Das geschah häufig in einem regulierenden Ton, z. B. durch Ermahnungen oder deutliche Gegenrede. Nur in einem Fall wurde ein Kommentar wegen Inzivilität gelöscht. Allerdings enthielten selbst überwiegend regulierende Moderationskommentare oft unterstützende Aspekte wie Fragen und wertschätzende Aussagen. Insgesamt entstand der Eindruck, als ob die Moderation keine Konfrontation herbeiführen wollte, sondern die Teilnehmer:innen auch bei Normverstößen dazu motivieren wollte, sich weiterhin zu beteiligen. Diese Aspekte entsprechen auch den teaminternen Zielen der Moderation, in einem respektvollen und wertschätzenden Ton Diskussionen anzuregen und die Teilnehmer:innen zu motivieren sowie auf die Einhaltung der Diskussionsregeln zu achten.

Ebenfalls eine zentrale Rolle spielten höfliche Umgangsformen, Danksagungen und Lob. Diese Komponenten zielten wahrscheinlich vor allem darauf ab, einen *respektvollen Umgangston* zu fördern. In einigen Fällen nahmen die Moderator:innen Bezug auf mehr als eine Person und sprachen weitere Diskussionsteilnehmer:innen an. Das könnte der *Reziprozität* zuträglich sein. Allerdings wurden Gespräche der Diskussionsteilnehmer:innen untereinander nur selten gefördert. Daher ist davon auszugehen, dass durch Bezugnahmen und Fragen der Moderation vor allem die Reziprozität zwischen Moderator:innen und Teilnehmer:innen gefördert wurde und nicht die Reziprozität zwischen Teilnehmer:innen untereinander. Der Versuch, weitere Teilnehmer:innen zu motivieren, ihre Meinung zu einem bestimmten Aspekt zu äußern, könnte außerdem die Meinungsvielfalt und damit die *Inklusivität* innerhalb der Diskussion fördern. Allerdings wurde diese Moderationskomponente nur vereinzelt eingesetzt. Qualitätskriterien jenseits klassischer Deliberation wie Storytelling, Emotionen und Humor wurden kaum gefördert. Diese Aspekte kamen in den Diskussionen aber ohnehin insgesamt selten vor. Mit Blick auf die Ergebnisse der Inhaltsanalyse zeigt sich hier allerdings auch, dass das Vorkommen von Qualitätsmerkmalen jenseits der Rationalität nicht durch das Erhalten von Moderationskommentaren honoriert wurde.

Insgesamt ist die Moderationsstrategie von #meinfernsehen2021 als förderlich für einen deliberativen Diskurs zu bewerten. Dabei zielte der Einsatz von Moderation insbesondere auf die Förderung von Rationalität sowie die Eindämmung von Inzivilität ab. Dies geschah hinsichtlich der Auswahl über alle Kommentare hinweg konsistent und spiegelt sich auch in den Moderationskommentaren wider. Allerdings hätte durch eine gezieltere Auswahl der moderationswürdigen Kommentare jenseits des klassisch rationalen Ideals eine höhere Sensibilität für die Erwünschtheit von anderen Dimensionen deliberativer Qualität, wie z. B. Reziprozität, expliziter Respekt und persönliche Erfahrungen, erreicht werden können. Mit Blick auf Inzivilität kann man von einer sehr konsequent zur Anwendung gebrachten Regulierung sprechen, es entsteht allerdings auch der Eindruck, dass die Moderation insgesamt sehr streng war und bereits bei milderen Verstößen regulierend eingriff, was laut Kritiker:innen der „klassischen" Deliberation weniger inklusiv ist (Bächtiger & Wyss, 2013; Papacharissi, 2004; Young, 2000). Das könnte von Teilnehmer:innen auch als „over-moderation" empfunden werden und kann im schlimmsten Fall einen offenen Diskurs behindern. Vereinzelt wurden Moderationsentscheidungen auch in der Diskussion kritisiert. Insgesamt wurde die Moderation von den Teilnehmer:innen der Evaluation jedoch überwiegend positiv bewertet und der Einsatz von Moderation mehrheitlich befürwortet (siehe Methodenteil in diesem Sammelband).

Die Studie leistet somit eine umfassende und systematische Bestandsaufnahme der Moderationsstrategie bei #meinfernsehen2021 und kann den Moderator:innen eine konsistente und weitgehend ihren Leitlinien entsprechende Umsetzung bescheinigen. Sie hat jedoch nicht die Untersuchung ihrer Wirkungen zum Gegenstand. Inwiefern Moderation demnach die Qualität von Diskussionsbeiträgen in diskursiven Online-Beteiligungsverfahren gezielt steigern kann, muss in einem weiteren Schritt untersucht werden. Einen ersten Anhaltspunkt bieten die Ergebnisse der Evaluation des Verfahrens durch die Teilnehmer:innen (siehe Methodenkapitel in diesem Sammelband): Hier zeige sich ein positiver Zusammenhang zwischen der Zufriedenheit mit der Moderation (Mittelwertindex, α =.96, M = 4.36, SD = 1.45) und der Zufriedenheit mit der Qualität der Diskussionen, $r(95)$ =.25, p =.02 sowie der Diskussionskultur, $r(95)$ =.33, p <.001. Künftige Studien sollten daher unterschiedliche Moderationsstrategien in verschiedenen Verfahren auf allen Stufen des Moderationsprozesses, a) der Auswahl des Moderationsobjekts und b) der moderativen Handlung vergleichen, um ein umfassenderes Bild zu Einsatz und Wirkungen von Moderation in diskursiven Online-Partizipationsverfahren zu bekommen.

Literatur

Bächtiger, André; & Wyss, Dominik (2013). Empirische Deliberationsforschung – eine systematische Übersicht. Zeitschrift Für Vergleichende Politikwissenschaft, 7(2), 155–181. https://doi.org/10.1007/s12286-013-0153-x

Blumler, Jay G.; & Coleman, Stephen (2001). Realising democracy online: A civic commons in cyberspace (Bd.. 2). London: IPPR.

Boberg, Svenja; Schatto-Eckrodt, Tim; Frischlich, Lena; & Quandt, Thorsten (2018). The moral gatekeeper? Moderation and deletion of user-generated content in a leading news forum. Media and Communication, 6(4), 58–69.

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

Edwards, Arthur R. (2002). The moderator as an emerging democratic intermediary: The role of the moderator in Internet discussions about public issues. Information Polity, 7(1), 3–20.

Einwiller, Sabine A.; & Kim, Sora (2020). How Online Content Providers Moderate User-Generated Content to Prevent Harmful Online Communication: An Analysis of Policies and Their Implementation. Policy & Internet, 12(2), 184–206.

Epstein, Dmitry; & Leshed, Gilly (2016). The magic sauce: Practices of facilitation in online policy deliberation. Journal of Deliberative Democracy, 12(1). https://doi.org/10.16997/jdd.244

Esau, Katharina; Frieß, Dennis; & Eilders, Christiane (2019). Online-Partizipation jenseits klassischer Deliberation: Eine Analyse zum Verhältnis unterschiedlicher Deliberationskonzepte in Nutzerkommentaren auf Facebook-Nachrichtenseiten und Beteiligungsplattformen. In: Ines Engelmann, Marie Legrand; & Hanna Marzinkowski (Hrg.), Digital Communication Research: Bd. 6. Politische Partizipation im Medienwandel (S. 221–245).

Frieß, Dennis; & Eilders, Christiane (2015). A systematic review of online deliberation research. Policy & Internet, 7(3), 319–339. https://doi.org/10.1002/poi3.95

Frieß, Dennis (2016). Online-Kommunikation im Lichte deliberativer Theorie: ein forschungsleitendes Modell zur Analyse von Online-Diskussionen. In: Philipp Henn; & Dennis Frieß (Hrsg.), Politische Online-Kommunikation: Voraussetzungen und Folgen des strukturellen Wandels der politischen Kommunikation (S. 143–169). Berlin https://doi.org/10.17174/dcr.v3.7

Frieß, Dennis; & Porten-Cheé, Pablo (2018). What Do Participants Take Away from Local eParticipation? Analyse & Kritik, 40(1), 1–29.

Grimmelmann, James (2015). The virtues of moderation. Yale Journal of Law & Technology, 17(1), 42–109.

Gutmann, Amy; & Thompson, Dennis F. (1996). Democracy and Disagreement. Cambridge, Mass.: Belknap Press.

Habermas, Jürgen (1983). Moralbewusstsein und kommunikatives Handeln. Suhrkamp.

Kuckartz, Udo (2018). Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung (4. Aufl.). Grundlagentexte Methoden. Beltz.

Masullo, Gina M.; Riedl, Martin J.; & Huang, Q. Elyse (2020). Engagement moderation: What journalists should say to improve online discussions. Journalism Practice, 1–17. https://doi.org/10.1080/17512786.2020.1808858

Muddiman, Ashley; & Stroud, Natalie J. (2017). News values, cognitive biases, and partisan incivility in comment sections. Journal of communication, 67(4), 586–609.

Newton, Kenneth (2012). Curing the Democratic Malaise with Democratic Innovations. In: Brigitte Geissel & Kenneth Newton (Hrsg.), Evaluating Democratic Innovations. Curing the Democratic Malaise? (S. 3–20), Routledge: New York.

Paasch-Colberg, Sünje; & Strippel, Christian (2021). „The Boundaries are Blurry …": How Comment Moderators in Germany See and Respond to Hate Comments. Journalism Studies. Vorab-Onlinepublikation. https://doi.org/10.1080/1461670X.2021.2017793

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

Smith, Graham (2009). Democratic innovations: Designing institutions for citizen participation. Cambridge University Press.

Steiner, Jürg (2012). The Foundations of Deliberative Democracy. Cambridge University Press. https://doi.org/10.1017/CBO9781139057486

Stroud, Natalie J.; Scacco, Joshua M.; Muddiman, Ashley; & Curry, Alexander L. (2015). Changing deliberative norms on news organizations' Facebook sites. Journal of Computer-Mediated Communication, 20(2), 188–203.

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

Wright, Scott; & Street, John (2007). Democracy, deliberation and design: the case of online discussion forums. New media & society, 9(5), 849–869.

Young, Iris M. (2002). Inclusion and Democracy. Oxford University Press.

Ziegele, Marc; & Jost, Pablo B. (2020). Not funny? The effects of factual versus sarcastic journalistic responses to uncivil user comments. Communication research, 47(6), 891–920.

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

Ziegele, Marc; Jost, Pablo; Frieß, Dennis; & Naab, Teresa K. (2019). Aufräumen im Trollhaus: zum Einfluss von Community-Managern und Aktionsgruppen in Kommentarspalten. Düsseldorf Institute for Internet and Democracy. Abrufbar unter: https://diid.hhu.de/wp-content/uploads/2019/04/DIID-Precis_Ziegele_V3.pdf; zuletzt abgerufen am 04.02.2022.

Ziegele, Marc; Quiring, Oliver; Esau, Katharina & Frieß, Dennis (2020). Linking News Value Theory With Online Deliberation: How News Factors and Illustration Factors in News Articles Affect the Deliberative Quality of User Discussions in SNS' Comment Sections. Communication Research, 47(6), 860–890. https://doi.org/10.1177/0093650218797884