



Précis [pʁe 'si:]

// APRIL 2019

Aufräumen im Trollhaus

Zum Einfluss von Community-Managern und Aktionsgruppen in Kommentarspalten

Marc Ziegele, Pablo Jost, Dennis Frieß & Teresa Naab

Kontakt: ziegele@hhu.de, Twitter: @ziegelem, @pbjost, @dennis_friess, @NaabTeresa

ABSTRACT

Theoretisch eröffnen Kommentarspalten einen Raum für sachliche, respektvolle und gleichberechtigte Diskussionen. Studien zeigen jedoch, dass Nutzerdiskussionen nur selten deliberative Normen wie Rationalität und Respekt erfüllen. Stattdessen enthalten sie oft Beleidigungen oder offene Aggressivität und Hassrede. Untersuchungen zeigen, dass diese inzivilen Kommentare eine Reihe von negativen Auswirkungen auf ihre Leser haben. Vor diesem Hintergrund stellt sich die Frage, wie die Qualität von Online-Diskussionen verbessert werden kann.

Das vorliegende Précis stellt zwei Studien vor, die sich dieser Frage nähern. Die erste Studie untersucht den Einfluss diskursiver journalistischer Moderation auf die Qualität von Nutzerkommentaren. Es zeigt sich u.a., dass ein vergemeinschaftender Stil respektvollere Diskussionen befördert. Die zweite Studie untersucht die Initiative #ichbinhier, welche sich zur Aufgabe gemacht hat, Online-Diskussionen zu versachlichen. Die Befunde zeigen, dass dieses Engagement die Qualität von Online-Diskussionen verbessern kann.



Dr. Marc Ziegele ist Jun.-Prof. an der Universität Düsseldorf und Leiter der Nachwuchsforschergruppe DEDIS

Pablo Jost, M.A. ist wissenschaftlicher Mitarbeiter an der Universität Mainz

Dennis Frieß, M.A. ist wissenschaftlicher Mitarbeiter an der Universität Düsseldorf

Dr. Teresa K. Naab ist akademische Rätin a. Z. an der Universität Augsburg

REFERENZEN ZUM THEMA

Esau, K., Frieß, D. & Eilders, C. (2017). Design Matters! An Empirical Analysis of Online Deliberation on Different News Plattformen. *Policy & Internet*, 9(3), 321-342.

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

Ziegele, M., & Jost, P. B. (2016). Not funny? The effects of factual versus sarcastic journalistic responses to uncivil user comments. *Communication Research*, online first, 1-30.

Einleitung

Das Kommentieren von Online-Nachrichten ist eine verbreitete Form politischer Anschlusskommunikation (Newman et al., 2016). Theoretisch eröffnen Kommentarspalten einen Raum für deliberative Öffentlichkeiten, in denen Nutzer politische Themen und Probleme sachlich, respektvoll und gleichberechtigt diskutieren können (Frieß & Eilders, 2015). Studien zeigen jedoch, dass Nutzerdiskussionen nur selten deliberative Normen wie Rationalität und Respekt erfüllen (Esau et al., 2017). Stattdessen enthalten sie oft inzivile, respektlose Äußerungen (Coe et al., 2014), also zum Beispiel Beleidigungen und Pöbeleien, aber auch offene Aggressivität und Hassrede. Derart inzivile Kommentare sind problematisch, da sie zu Meinungspolarisierung führen (Anderson et al., 2014), Vorurteile fördern (Hsueh et al., 2015), prosoziales Verhalten gegenüber hilfsbedürftigen Minderheiten verringern (Ziegele et al., 2018b) und die wahrgenommene Qualität von Nachrichtenartikeln beeinträchtigen können (Prochazka et al., 2018). Auch viele potenzielle Diskussionsteilnehmer*innen fühlen sich durch den rauen Ton in Online-Kommentarspalten abgeschreckt (Springer et al., 2015).

Vor diesem Hintergrund stellt sich die Frage, wie die Qualität von Online-Diskussionen verbessert werden kann. Medienunternehmen filtern oder löschen inzivile Kommentaren häufig manuell oder mithilfe von Wortlisten. Allerdings sind derartige Formen der Diskussionsmoderation oft intransparent und fehleranfällig und daher problematisch, wenn Diskussionen deliberativ gestaltet werden sollen. Eine vielversprechendere und gleichberechtigtere Form der Diskussionsmoderation ist die sogenannte diskursive Moderation (Ziegele & Jost, 2016): Hierbei begegnen Community-Manager*innen von Nachrichten-Medien respektlosen, inzivilen und problematischen Beiträgen unter anderem mit Gegenrede, dem sogenannten Counter Speech (Kramp & Weichert, 2018). Darüber versuchen sie, die Autor*innen der problematischen Kommentare, aber auch stille Leser*innen zum Verfassen von zivilen und

sachlichen Kommentaren anzuregen. Doch kann dieses Engagement nachhaltig die Qualität der Diskussionen in Kommentarspalten verbessern? Bislang gab es auf diese Fragen nur vorläufige Antworten: Eine Case Study von Leif Kramp und Stephan Weichert (2018) sowie ein Laborexperiment von Marc Ziegele und Pablo Jost (2016) lieferten erste Hinweise darauf, dass sich das diskursive Engagement von Community-Manager*innen in Kommentarspalten lohnen kann. Eine Bestätigung dieser Befunde auf einer breiten Datenbasis stand allerdings bislang aus.

Studie 1: Wirkungen diskursiver Moderation durch Community-Manager*innen

Wir haben daher zunächst untersucht, ob diskursive Moderation von problematischen Kommentaren durch Community-Manager*innen, in Form von Gegenrede, die Qualität von nachfolgenden Kommentaren erhöht (vgl. Ziegele et al., 2018a). Dafür haben wir zunächst in einer qualitativen Studie vier Stile identifiziert, wie Community-Manager*innen von großen Nachrichtenmedien auf die Kommentare ihrer Nutzer antworten. Diese Stile sind in Abbildung 1 dargestellt.

Die Systematisierung unterscheidet zum einen, ob Moderator*innen eher ermutigendes oder regulierendes Feedback auf die Kommentare ihrer Nutzer*innen geben. Bei ermutigendem Feedback zeigen die Moderator*innen Interesse daran, mit den Nutzer*innen inhaltlich zu diskutieren (*Diskutieren*). Sie bitten die Verfasser*innen der Kommentare zum Beispiel, zusätzliche Argumente zu liefern, fügen eigene Argumente hinzu oder stellen Zusatzinformationen bereit. Bei der zweiten Form des ermutigenden Feedbacks (*Vergemeinschaften*) versuchen Moderator*innen, eine informelle und angenehme Gesprächsatmosphäre zu schaffen. Sie zeigen Verständnis für die Gefühle der Kommentator*innen, weisen auf Gemeinsamkeiten zwischen verschiedenen Positionen hin oder bringen erleichternden Humor ein. Bei den regulierenden Formen des Feedbacks versuchen Moderator*innen zum einen, Nutzer*innen direkt auf Fehlverhalten aufmerksam zu ma-

chen bzw. sie zu bitten, respektvoller miteinander umzugehen (*Erziehen*). Zum anderen nutzen Moderator*innen zum Teil offensive Methoden, z.B. Sarkasmus und/oder Ironie, um Fehlverhalten von Nutzer*innen aufzudecken bzw. um den Verfasser*innen der Kommentare einen Spiegel vorzuhalten (*Konfrontieren*).

wir die Qualität von SLK, die im Anschluss an *unmoderierte* Top-Level-Kommentare gepostet wurden, mit der Qualität von SLK vergleichen, die im Anschluss an *moderierte* Top-Level-Kommentare gepostet wurden.

		Art des Feedbacks	
		Ermutigend	Regulierend
Art der Moderation	Deliberativ	Diskutieren Inhaltliche Argumente hinzufügen; Fragen klären oder stellen; Zusatzwissen bereitstellen; sachlich mit Nutzer*innen diskutieren	 Erziehen Nutzer*innen bitten, sich respektvoll zu verhalten; auf Verletzungen der Netiquette hinweisen
	Nicht-deliberativ	Vergemeinschaften Informelle und angenehme Diskussionsatmosphäre schaffen; Verständnis für Gefühle der Kommentierenden zeigen; „Witzeln“; Gemeinsamkeiten aufzeigen	 Konfrontieren Kommentare offensiv und/oder sarkastisch angreifen; Autor*innen von Kommentaren bloßstellen; „Gegen-Trollen“

Abbildung 1: Identifizierte Moderationsstile der Community-Manager*innen von Nachrichtenseiten

Um zu untersuchen, ob diese vier Moderationsstile die Qualität der Online-Diskussionen in Kommentarspalten verbessern können, haben wir eine quantitative Inhaltsanalyse durchgeführt. Im Februar und Oktober 2016 haben wir alle Artikel (N=10.081) und Nutzerkommentare (N=2.103.076) sowie die Moderationskommentare von Community-Manager*innen (N=8.497) heruntergeladen, die auf den Facebook-Seiten von 15 deutschen Nachrichtenmedien (u.a. Tagesschau.de, Spiegel Online, Die Welt, FAZ) veröffentlicht wurden.

Aus dieser Grundgesamtheit haben wir eine ausgewogene Stichprobe von insgesamt 3.206 Top-Level-Kommentaren (TLK, d.h. Kommentare, die direkt auf den Artikel antworten), 4.901 Sub-Level-Kommentaren (SLK, d.h. Kommentare, die auf andere Kommentare antworten) und 1.656 Moderations-SLK (d.h. Kommentare von Community-Manager*innen, die auf andere Kommentare antworten) gezogen. Dadurch konnten

Einundfünfzig geschulte studentische Hilfskräfte codierten die insgesamt 9.763 Kommentare hinsichtlich folgender, aus der Deliberationsforschung abgeleiteter Qualitäts-Kriterien:

- *Respekt*, mit den Sub-Dimensionen *Höflichkeit* (zum Beispiel Abwesenheit von Beleidigungen und Vulgärsprache) und *Zivilität* (zum Beispiel Abwesenheit von negativen Stereotypen sowie Abwesenheit von Rassismus und Sexismus)
- *Diskursivität*, zum Beispiel die Präsenz von Argumenten, Zusatzwissen und konstruktiven Lösungsvorschlägen.

Darüber hinaus wurden die Moderations-Kommentare hinsichtlich der Präsenz eines oder mehrerer der oben beschriebenen vier Moderations-Stile *Diskutieren*, *Regulieren*, *Vergemeinschaften* und/oder *Konfrontieren* codiert.

Um die Frage nach den potenziellen Wirkungen von diskursiver Moderation zu beantworten, analysierten wir, ob die Antwort-Kommentare (SLK)

auf problematische Top-Level-Kommentare (TLK) *respektvoller* (d.h. weniger unhöflich und inzivil) und *diskursiver* (d.h. rationaler und konstruktiver) sind, wenn auf die TLK zuvor bereits ein Moderator oder eine Moderatorin geantwortet hatte.¹ Die Vergleichsgruppe bildeten problematische TLK, auf die keine Moderator*innen geantwortet hatten. In der Gruppe der moderierten problematischen TLK untersuchten wir immer nur die SLK, die nach dem Moderationskommentar erschienen waren. In unseren statistischen Mehrebenen-Modellen kontrollierten wir zusätzlich verschiedene Drittvariablen (z.B. das Medium, bei dem der Kommentar gepostet wurde und das Thema des zugehörigen Nachrichtenartikels). Dadurch erhöhten wir die Wahrscheinlichkeit, dass etwaige Unterschiede in der Qualität der Folgediskussionen tatsächlich auf die verschiedenen Moderationsstile zurückzuführen sind. Die zusammengefassten Ergebnisse unserer Analyse sind in Abbildung 2 dargestellt. Die vollständige Ergebnistabelle findet sich in Anhang A-1.

schen verschiedenen Positionen aufzeigten oder „harmlosen“ Humor einsetzten, erschienen danach weniger respektlose Kommentare, als wenn der problematische Kommentar unmoderiert geblieben wäre. Der Moderationsstil *Erziehen* verschlechterte dagegen das Ausmaß an Respekt in den Anschlussdiskussionen; wenn Moderator*innen also versuchten, ihre Nutzer*innen mit dem „erhobenen Zeigefinger“ zu einer respektvolleren Diskussion zu bewegen, hat dies unseren Daten zufolge eher ins Gegenteil umgeschlagen. Keine signifikanten Verbesserungen oder Verschlechterungen des Ausmaßes an Respekt in den Anschlussdiskussionen zogen die Moderationsstile *Diskutieren* und *Konfrontieren* nach sich.

Darüber hinaus untersuchten wir die möglichen Auswirkungen der vier Moderationsstile auf die *Diskursivität* der Anschlussdiskussionen, d.h. auf das Ausmaß an argumentgeführter und konstruktiver Debatte.

Moderationsstil	Zusammenhänge mit...	
	Respekt in Anschlussdiskussionen (Zivilität/Höflichkeit)	Diskursivität der Anschlussdiskussionen (Rationalität/Konstruktivität)
 Diskutieren		
 Erziehen		
 Vergemeinschaften		
 Konfrontieren		

Abbildung 2: Grüne/rote Pfeile bedeuten signifikant positive/negative Zusammenhänge. Graue Pfeile beschreiben Tendenzen, die jedoch keine statistische Signifikanz erreichen.

Zunächst zum Ausmaß an *Respekt* in der Anschlussdiskussionen, d.h. der Abwesenheit von Beleidigungen, Pöbeleien und Hass. Unseren Analysen zufolge waren die Anschlussdiskussionen unter Kommentaren, die mit dem Stil *Vergemeinschaften* moderiert wurden, respektvoller als die Anschlussdiskussion unter vergleichbaren nicht-moderierten Kommentaren. Das heißt: Wenn Moderator*innen problematischen Kommentaren mit Empathie begegneten, Gemeinsamkeiten zwi-

Die Analysen zeigen, dass der Moderationsstil *Diskutieren* die Diskursivität der Anschlussdiskussionen signifikant erhöhte. Wenn Moderator*innen selbst Argumente oder Zusatzwissen einbrachten und damit „vorlebten“, wie man sachlich diskutieren kann, haben Kommentator*innen sich scheinbar daran orientiert. Eine geringere Diskursivität der Anschlussdiskussionen ging dagegen mit den Moderationsstilen *Vergemeinschaften* und *Konfrontieren* einher.

Dies unterstreicht, dass das ironisch-sarkastische Bloßstellen von Kommentator*innen, die sich unangemessen verhalten, aus diskursiver Perspektive keine nachhaltig gewinnbringende Lösung zu sein scheint. Die verringerte Diskursivität der Anschlussdiskussionen nach vergemeinschaftender Moderation kann vermutlich auf einen gestiegenen Anteil an Small-Talk und anderen nicht-deliberativen Kommunikationsformen zurückgeführt werden. Je nach Ziel der Moderation und intendierter Funktion des Diskussionsforums müssen Moderator*innen also abwägen, ob sie selbst durch einen diskursiven Stil zu einer argumentativen und konstruktiven Debatte beitragen möchten oder durch einen stärkeren Fokus auf kohäsionsstiftende Elemente Höflichkeit und Zivilität unter den Diskutant*innen befördern möchten. Auch der Einsatz von konfrontativer Moderation ist, je nach Zielsetzung, durchaus vorstellbar – immerhin konnte eine frühere Studie zeigen, dass das ironisch-sarkastische „Zurücktrollen“ durch Moderator*innen den Unterhaltungswert der Diskussionen erhöhen kann (Ziegele & Jost, 2016).

Studie 2: Qualität und Wirkung diskursiver Moderation durch die Aktionsgruppe #ichbinhier

Heutzutage versuchen nicht nur professionelle Moderator*innen, die Ordnung in Kommentarspalten wiederherzustellen bzw. die Qualität von Online-Diskussionen zu verbessern. Immer mehr Nutzer*innen versammeln sich in sogenannten Aktionsgruppen, um gemeinsam dem Hass und der Hetze in Online-Diskussionen entgegenzutreten. Hier entsteht eine neue Form der digitalen Zivilcourage, die bisher noch wenig beforscht wurde. Die in Deutschland größte Aktionsgruppe ist *#ichbinhier*. Die nach schwedischem Vorbild gegründete Gruppe hat aktuell rund 45.000 Mitglieder, die sich in einer gleichnamigen Facebook-Gruppe koordinieren und mit sogenannten Aktionen gezielt in inzivile, respektlose und eskalierte Diskussionen auf den Facebook-Seiten von Nachrichtenmedien wie Tagesschau und Focus Online eingreifen. Ziel des Engagements ist eine Versachlichung der Debatten durch rationale und

respektvolle Gegenrede. Für die Gestaltung ihrer Gegenrede werden den #ichbinhier-Mitgliedern Empfehlungen gegeben, die deliberative Normen wie Respekt und Sachlichkeit aufgreifen. So heißt es zum Beispiel: „Sei empathisch und respektvoll, untermauere deine Aussagen mit Fakten und Argumenten, belehre andere Diskussionsteilnehmer nicht, bleibe beim Thema, lasse dich nicht auf Provokationen ein“ (Ley, 2018).

Doch sind die Kommentare von #ichbinhier-Mitgliedern tatsächlich deliberativer bzw. qualitativ „hochwertiger“ als die Kommentare von anderen Nutzer*innen? Und hat die Gegenrede der Mitglieder von #ichbinhier einen messbaren positiven Effekt auf die Qualität der Anschlussdiskussionen? Um diese Fragen zu beantworten, haben wir eine zweite Studie durchgeführt. Ziel war es, Aussagen darüber zu treffen, inwieweit zivilgesellschaftliches Engagement professionelle Moderation unterstützen und darüber die Qualität der Debatten in Kommentarspalten verbessern kann.

Für eine quantitative Inhaltsanalyse stellte uns #ichbinhier einen Datensatz zur Verfügung, der alle Aktionen der Gruppe zwischen November 2017 und Januar 2018 enthielt.² Insgesamt lagen N=14.104 Top-Level-Kommentare (TLK) von 21 Facebook-Seiten deutscher Nachrichtenmedien (z.B. Tagesschau, FOCUS Online, FAZ.net, ZDF heute) sowie die ersten fünf Antwort-Kommentare (SLK) zu jedem TLK vor. Für die Analyse wurde, analog zu Studie 1, aus dieser Grundgesamtheit eine balancierte Stichprobe gezogen: Für jedes Nachrichtenmedium wurden zunächst zufällig sechs Nachrichtenartikel ausgewählt. Zu jedem Artikel wählten wir dann zufällig sechs TLK von #ibh-Mitgliedern, sechs TLK von Nicht-Mitgliedern sowie die den TLK zugeordneten ersten fünf SLK aus. Der finale Datensatz beinhaltete n=641 TLK und n=2.928 SLK. Die Identifikation der Mitglieder von #ichbinhier im Datensatz erfolgte unter anderem über einen Abgleich der Namen der Kommentarverfasser*innen mit der Mitgliederliste der Gruppe. Dadurch konnten #ichbinhier-Mitglieder auch dann identifiziert werden, wenn sie ohne den #ichbinhier-Hashtag Kommentare verfassten.

Sechs geschulte studentische Hilfskräfte codierten für alle TLK und SLK die An- oder Abwesenheit von insgesamt 27 Qualitäts-Kriterien aus den Dimensionen Diskursivität (Rationalität/Konstruktivität, darunter Argumente, Zusatzwissen, Lösungsvorschläge, Fragen, usw.) und Respekt (Zivilität/Höflichkeit, darunter Beleidigungen, Rassismus, Stereotype, usw.).³ Für die Auswertung verglichen wir den Anteil der Kommentare (TLK und SLK) von #ichbinhier-Mitgliedern, der bestimmte Qualitäts-Kriterien erfüllte, mit dem Anteil der Kommentare (TLK und SLK) von Nicht-Mitgliedern, der diese Kriterien erfüllte. Abbildung 3 zeigt beispielhaft, in wie viel Prozent der Kommentare von #ichbinhier-Mitgliedern und Nicht-Mitgliedern das jeweilige Qualitäts-Kriterium zu finden war. So enthielten zum Beispiel 50% der Kommentare von #ichbinhier-Mitgliedern Argumente, während dies nur auf 36% der Kommentare von Nicht-Mitgliedern zutraf. Weiterhin waren die Kommentare von #ichbinhier-Mitglie-

wie Nicht-Mitglieder (16% vs. 14% bei Nicht-Mitgliedern). Insgesamt zeigt die Tabelle im Anhang A-2, dass die deliberative Qualität der Kommentare von #ichbinhier-Mitgliedern über alle vier erhobenen Qualitäts-Dimensionen (Rationalität, Konstruktivität, Zivilität und Höflichkeit) höher war als die Qualität der Kommentare von Nicht-Mitgliedern.⁴ Statistisch gesehen tragen die Kommentare von #ichbinhier insofern allein durch ihre Präsenz dazu bei, dass das durchschnittliche Niveau von Online-Diskussionen steigt.

Doch motivieren die Kommentare der #ichbinhier-Mitglieder andere Nutzer*innen auch dazu, bessere Anschlussdiskussionen in den Antwort-Kommentaren (SLK) zu führen? Um dies zu prüfen, haben wir mittels Mehrebenenanalysen untersucht, ob die SLK unter #ichbinhier Top-Level-Kommentaren (TLK) respektvoller (d.h. weniger unhöflich und inzivil) und diskursiver (d.h. rationaler und konstruktiver) sind als die SLK unter

	 Argumente	 Themenbezug	 Lösungsvorschläge	 Geringschätzungen	 Negative Emotionen	 Stereotype
Kommentare von #ichbinhier-Mitgliedern	50%	70%	20%	11%	16%	3%
Kommentare von anderen Nutzern	36%	53%	6%	32%	14%	11%

Abbildung 3: Anteile der Kommentare von #ichbinhier-Mitgliedern und anderen Nutzer*innen, die verschiedene deliberative Qualitäts-Kriterien enthalten. Basis: 3.562 Kommentare (TLK und SLK).

dern signifikant häufiger themenbezogen (70% vs. 53% bei Nicht-Mitgliedern) und sie enthielten häufiger konstruktive Lösungsvorschläge (20% vs. 6% bei Nicht-Mitgliedern). Hinsichtlich der Negativ-Kriterien Inzivilität und Unhöflichkeit enthielten Kommentare von #ichbinhier-Mitgliedern signifikant seltener Geringschätzungen (11% vs. 32% bei Nicht-Mitgliedern) und antagonistische Stereotype (3% vs. 11% bei Nicht-Mitgliedern). Allerdings äußerten #ichbinhier-Mitglieder in ihren Kommentaren ähnlich häufig negative Emotionen

TLK, die von Nicht-Mitgliedern verfasst wurden. In unseren Modellen haben wir verschiedene Drittvariablen kontrolliert, darunter das Medium, bei dem der Kommentar gepostet wurde und das Thema des Artikels. Zusätzlich kontrollierten wir, ob ein SLK von einem #ichbinhier-Mitglied geschrieben wurde – denn wenn bessere Anschlussdiskussionen an TLK Kommunikation an TLK von #ichbinhier-Mitgliedern nur dadurch zustande kommen, dass #ichbinhier-Mitglieder auf diese Kommentare antworten, wäre keine „echte“ Ver-



Abbildung 4: Respekt und Diskursivität der Anschlusskommunikation und Sub-Level-Kommentaren (SLK) in Abhängigkeit von Eigenschaften der Top-Level-Kommentare (Diskursivität, Respekt, #ichbinhier-Mitgliedschaft). Grüne Pfeile bedeuten signifikant positive Zusammenhänge. Graue Pfeile beschreiben Tendenzen, die jedoch keine statistische Signifikanz haben.

besserung erreicht. Die Ergebnisse zeigen, dass die Anschlussdiskussionen zu #ichbinhier-Kommentaren tatsächlich signifikant respektvoller und diskursiver sind, wenn nur die #ichbinhier-Mitgliedschaft auf TLK-Ebene, nicht aber die Kontrollvariablen, als Prädiktoren berücksichtigt werden. Weiterführende Analysen unter Einbezug der Kontrollvariablen zeigen jedoch, dass dies vor allem daran liegt, dass andere #ichbinhier-Mitglieder auf die #ichbinhier-TLK antworten und diese Antworten signifikant respektvoller und diskursiver ausfallen (vollständige Ergebnistabellen im Anhang A-3). Im finalen Modell bleiben die in Abbildung 4 dargestellten Zusammenhänge signifikant. Demnach sind Antwort-Kommentare auf Top-Level-Kommentare (TLK) signifikant respektvoller (d.h. höflich und zivil), wenn die jeweiligen TLK ebenfalls respektvoll sind. Signifikant diskursiver sind die Antwort-Kommentare auf TLK, wenn die jeweiligen TLK ebenfalls diskursiv sind. Respektvolle und diskursive Kommentare scheinen demnach tatsächlich qualitativ hochwertige Anschlussdiskussionen zu stimulieren, und zwar unabhängig davon, ob diese Kommentare von #ichbinhier-Mitgliedern oder „normalen“ Nutzer*innen geschrieben sind. Da unsere Ergebnisse jedoch auch gezeigt haben, dass #ichbinhier-Mitglieder signifikant häufiger als andere Nutzer*innen respektvolle und diskursive Kommentare schreiben (siehe Abbildung 3), ist ein indirekter positiver Effekt der #ichbinhier-Mitgliedschaft auf die Qualität der Anschlussdiskussionen denkbar.

Fazit

Eingangs haben wir die Frage gestellt, ob das Engagement von Community-Manager*innen und zivilgesellschaftlichen Aktionsgruppen für eine bessere Diskussionskultur in Online-Kommentarspalten eine echte Veränderung bringt oder ob es ein Kampf gegen Windmühlen bleibt. Unsere Studien liefern starke empirische Hinweise darauf, dass wir Kommentarspalten nicht sich selbst überlassen sollten: Community-Manager*innen können, wenn sie einen respektvollen, vergemeinschaftenden Ton treffen und sachlich mitdiskutieren, sehr wohl die Qualität der anschließenden Diskussionen erhöhen. Nicht uneingeschränkt empfehlenswert ist dagegen ein „erzieherischer“ Moderationsstil, der sich darauf konzentriert, Diskussionsteilnehmer*innen ihr Fehlverhalten „mit dem Zeigefinger“ aufzuzeigen sowie ein konfrontativer Moderationsstil, der Nutzer*innen potenziell bloßstellt.

Auch die überwiegend respektvollen und diskursiven Kommentare der Aktionsgruppe #ichbinhier können dazu beitragen, dass stille Leser*innen ein besseres Gesamt-Diskussionsklima wahrnehmen. Da experimentell gezeigt wurde, dass eine derartige Wahrnehmung die Partizipationsbereitschaft der Leser*innen erhöht (Ziegele & Jost, 2016), könnten Initiativen wie #ichbinhier langfristig eine vielversprechende Ergänzung zu professioneller Moderation sein und deliberative Partizipation in

Kommentarspalten fördern. Ähnlich wie Moderator*innen, die nur über den passenden Moderationsstil eine Veränderung bewirken können, führt die #ichbinhier-Mitgliedschaft allerdings nicht automatisch dazu, dass sich die unmittelbar an die Kommentare der Mitglieder anschließenden Diskussionen verbessern; dies geschieht vor allem dann, wenn die Kommentare selbst deliberative Kriterien wie Respekt und Diskursivität erfüllen. Das ist zwar bei Kommentaren von #ichbinhier-Mitgliedern häufig der Fall. Dennoch konnten wir in Elf Prozent der #ichbinhier-Kommentare Geringschätzungen anderer Personen, Ideologien oder Ideen feststellen. Dieser Anteil ist sogar noch etwas höher (knapp unter 20%), wenn #ichbinhier-Mitglieder ohne Verwendung des #ichbinhier-Hashtags kommentieren. Auch wenn die #ichbinhier-Mitglieder die Gründe für ihre Geringschätzungen häufig ausführlich benennen, würde es dem Respekt und der Diskursivität in Diskussionen nützen, die eigenen Überzeugungen nicht für unverrückbar richtig zu halten und zu respektieren, dass und warum Diskussions-„Gegner“ auch andere Ansichten haben.

Unsere Studien weisen diverse Limitationen auf. Mit Inhaltsanalysen können keine kausalen Zusammenhänge aufgezeigt werden, da immer Drittvariablen existieren, die wir nicht kontrollieren können, die aber ebenfalls Einfluss auf die Ergebnisse nehmen können. Unsere Befunde sollten daher in kontrollierten Experimenten repliziert werden. Darüber hinaus können die gewählten Untersuchungszeiträume sowie das verfügbare Datenmaterial der Studien nur eine Momentaufnahme der Situation in Online-Kommentarspalten zu einer bestimmten Zeit und auf bestimmten Nachrichtenseiten geben. Die Studien sind daher nur ein erster Schritt für ein umfangreiches Forschungsprogramm, das Community-Manager*innen, aber auch Mitgliedern von Aktionsgruppen sowie individuellen Nutzer*innen praktische Tipps geben kann, wie sie dazu beitragen können, Online-Diskussionen aktiv in einem demokratischen Sinne zu gestalten. Dass sich das Engagement lohnt, dafür liefern unsere Studien erste wertvolle Hinweise.

Acknowledgement

Die Veröffentlichung wurde im Rahmen des Forschungsverbundes NRW Digitale Gesellschaft durch das Ministerium für Kultur und Wissenschaft des Landes Nordrhein-Westfalen gefördert.

Fußnoten

¹ In der Studie ging es vor allem um die Frage, ob diskursive Moderation die Folgediskussionen zu problematischen, d.h. inzivilen und/oder unhöflichen Top-Level-Kommentaren verbessern kann. Daher werden hier nur TLK betrachtet, die entweder inzivil oder unhöflich waren (n = 1.464).

² Für ihre Unterstützung danken wir vor allem Philip Kreißel und Susanne Tannert von #ichbinhier.

³ Für ihre Unterstützung danken wir Johannes Binder, Vivien Hantusch, Lukas Müller, Jana Peters und Sarah Reuter vielmals.

⁴ Ein detaillierter Vergleich aller erhobenen Qualitäts-Kriterien kann auf Anfrage zur Verfügung gestellt werden.

Literatur

- 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, 373–387.
- Coe, K., Kenski, K., & Rains, S. A. (2014). Online and uncivil? Patterns and determinants of incivility in newspaper website comments. *Journal of Communication*, 64, 658–679.
- Esau, K., Frieß, D. & Eilders, C. (2017). Design Matters! An Empirical Analysis of Online Deliberation on Different News Platforms. *Policy & Internet*, 9(3), 321-342.
- Friess, D., & Eilders, C. (2015). A systematic review of online deliberation research. *Policy & Internet*, 7, 319–339.
- 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, 557–576.

- Kramp, L., & Weichert, S. (2018). Hass im Netz: Steuerungsstrategien für Redaktionen. *Schriftenreihe Medienforschung der LfM (80)*. Leipzig: VISTAS.
- Ley, H. (2018). *#ichbinhier: Zusammen gegen Fake News und Hass im Netz*. Köln: DuMont.
- Newman, N., Fletcher, R., Kalogeropoulos, A., Levy, D. A. L., & Nielsen, R. K. (2016). *Reuters Institute Digital News Report 2016*. Online verfügbar unter <http://www.digitalnewsreport.org/>
- Prochazka, F., Weber, P., & Schweiger, W. (2018). Effects of civility and reasoning in user comments on perceived journalistic quality. *Journalism Studies*, 19, 62–78.
- Ziegele, M., & Jost, P. B. (2016). Not funny? The effects of factual versus sarcastic journalistic responses to uncivil user comments. *Communication Research*, online first, 1–30.
- Ziegele, M., Jost, P., Bormann, M., & Heinbach, D. (2018a). Journalistic counter-voices in comment sections: Patterns, determinants, and potential consequences of interactive moderation of uncivil user comments. *SCM Studies in Communication and Media*, 7(4), 525-554.
- Ziegele, M., Koehler, C., & Weber, M. (2018b). Socially Destructive? Effects of Negative and Hateful User Comments on Readers' Donation Behavior toward Refugees and Homeless Persons. *Journal of Broadcasting & Electronic Media*, 62(4), 636-653.

Anhang

Anhang A-1 – Ergebnisse der Mehrebenenanalyse für die Effekte von journalistischer Moderation auf den Respekt in und die Diskursivität der Anschlussdiskussionen

	Modell 1: Respekt in Anschlussdiskussion	Modell 2: Diskursivität der Anschlussdiskussionen
	<i>B (SE)</i>	<i>B (SE)</i>
Eigenschaften der TLK		
Höflichkeit	0.89*** (0.14)	0.06 (0.10)
Zivilität	0.80*** (0.12)	0.22 (0.12)
Stil der Moderation s-SLK		
Diskutierender Stil	-0.15 (0.12)	0.36*** (0.10)
Erziehender Stil	-0.41* (0.17)	0.08 (0.15)
Konfrontativer Stil	-0.20 (0.18)	-0.29* (0.15)
Vergemeinschaftender Stil	0.39** (0.14)	-0.71*** (0.11)
Kontrollvariablen		
n-tv	0.01 (0.27)	-0.57** (0.22)
Sueddeutsche Zeitung	-0.12 (0.28)	-0.18 (0.21)
Frankfurter Allgemeine Zeitung	-0.41 (0.33)	0.01 (0.24)
Berliner Morgenpost	-0.91* (0.40)	-0.62* (0.31)
Deutschlandfunk	0.10 (0.43)	0.33 (0.34)
BR24	-0.13 (0.27)	-0.16 (0.21)
HR-info	-0.17 (0.28)	0.66** (0.23)
Krautreporter	-1.19* (0.33)	0.33 (0.25)
Die Welt	-0.20 (0.26)	-0.35 (0.20)
ZDF heute	0.13 (0.31)	-0.04 (0.24)
Hannoversche Allgemeine	-0.47 (0.31)	-0.18 (0.25)
Tagesspiegel	-0.24 (0.35)	-0.00 (0.27)
Spiegel Online	-0.14 (0.27)	-0.36 (0.23)
RTLII News	-0.67 (0.54)	-0.69 (0.51)
ARD Tagesschau (Ref Cat.)	1	
<i>n</i> TLK (<i>n</i> SLK)	1.461 (3.929)	1.461 (3.929)

Anmerkungen: * $p < .05$, ** $p < .01$, *** $p < .001$

Anhang A-2: Vergleich der Eigenschaften der Top-Level- und Sub-Level-Kommentare von #ibh-Mitgliedern und Nicht-Mitgliedern

	Kommentare							
	#ichbinhier-Mitglieder			Nicht-Mitglieder			<i>t</i>	<i>df</i>
	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>		
Rationalität ^{a)}	1.17	1.26	1737	0.80	1.08	1737	-14.16***	3416.29
Konstruktivität ^{b)}	1.12	0.86	1737	0.74	0.73	1832	-9.61***	3403.28
Inzivilität ^{c)}	0.04	0.20	1737	0.21	0.54	1832	12.50***	2343.71
Unhöflichkeit ^{d)}	0.24	0.58	1737	0.70	0.93	1832	17.78***	3072.36

Anmerkungen: *N* = 3.569 Nutzerkommentare (TLK und SLK); * $p < .05$, ** $p < .01$, *** $p < .001$

a) Summenindex (Werte von 0-5) aus den Sub-Dimensionen Argumente, rationale Argumente, moralische Argumente, ausgewogene Argumente, persönliche Erfahrungen (jeweils dichotom codiert).

b) Summenindex (Werte von 0-3) aus den Sub-Dimensionen Deliberative Lösungsvorschläge, Zusatzwissen, Appelle, Fragen (jeweils dichotom codiert).

c) Summenindex (Werte von 0-6) aus den Sub-Dimensionen Dehumanisierung, Stereotype, Sexismus, Rassismus, Lügenvorwürfe, Drohungen (jeweils dichotom codiert).

d) Summenindex (Werte von 0-5) aus den Sub-Dimensionen Schreien, Beleidigung, Sarkasmus/Zynismus, Vulgarität, Geringschätzungen (jeweils dichotom codiert).

Anhang A-3: Ergebnisse der Mehrebenen-Regression für die Effekte von #ichbinhier-Interventionen in TLK auf den Respekt in und die Diskursivität von SLK

	AV: Respekt in SLK		AV: Diskursivität in SLK	
	Modell 1: nur #ichbinhier	Modell 2: alle Prädiktoren	Modell 1: nur #ichbinhier	Modell 2: alle Prädiktoren
	<i>Koeffizient</i>	<i>Koeffizient</i>	<i>Koeffizient</i>	<i>Koeffizient</i>
(Konstante)	0.06 ^{***}	0.07 ^{***}	0.16 ^{***}	0.12 ^{***}
<i>TLK-Prädiktoren</i>				
#ibh-TLK (1 = ja)	0.02 ^{***}	0.00	0.03 ^{***}	-0.01
Respekt im TLK	.	0.04 [*]	.	-0.05
Diskursivität des TLK	.	-0.01	.	0.18 ^{***}
<i>SLK-Prädiktoren</i>				
#ibh-SLK (1 = ja)	.	0.05 ^{***}	.	0.05 ^{***}
Zufällige Effekte (σ^2)	0.01	0.01	0.02	0.02
ICC				
TLK-Ebene	0.11	0.12	0.25	0.22
Artikel-Ebene	0.02	0.02	0.06	0.04
Medien-Ebene	0.00	0.01	0.00	0.00
Marginal R ²	0.02	0.09	0.01	0.07
AIC	-5.715,17	-5.913,54	-2.871,14	-2.976,70
n (SLK / TLK / Artikel / Medien)	2.807 / 602 / 59 / 10			

Anmerkungen. TLK = Top-Level-Kommentar, SLK = Sub-Level-Kommentar (Antwort-Kommentar);

* $p < .05$, ** $p < .01$, *** $p < .001$